Cross-modal, auditory-visual Stroop interference and possible implications for speech memory

NELSON COWAN and ALEXANDER BARRON
University of Missouri, Columbia, Missouri

This study examines effects of auditory color-word interference on a visual Stroop task with a spoken response. The presence of cross-modal interference indicates that subjects could not prevent the processing of irrelevant, spoken color words. Additional aspects of the results (e.g., lack of effects from noncolor items and additivity of auditory and visual interference) are used to support a description of processing in which multiple verbal items enter a prespeech buffer and a selection mechanism examines buffer items in parallel.

The interrelated concepts of *buffer memory* and *selective attention*, introduced into cognitive psychology nearly 30 years ago (e.g., Broadbent, 1958), are still central to contemporary cognitive theory. Nevertheless, a great deal of uncertainty remains about their characteristics. We hope to clarify certain basic characteristics of memory storage and selectivity through a study of irrelevant speech input in a cross-modal color-word interference (Stroop, 1935) task. Below, questions about a prespeech buffer and selective attention will be discussed, and then the relevance of a cross-modal Stroop procedure will be explained.

A Prespeech Buffer Memory

The term "buffer memory" will simply refer to the temporary products of a particular type of information processing, available for additional processing if the appropriate resources are applied before the information is lost. This definition avoids unnecessary assumptions about the characteristics of buffers. It is not necessary to specify whether the buffer is a static representation or a dynamic byproduct of processing (cf. Cowan, 1984, pp. 363-364). A static representation is usually associated with a multistore model (Atkinson & Shiffrin, 1968) and a dynamic representation with the levels-of-processing approach (Craik & Lockhart, 1972), but, as Monsell (1984, p. 329) pointed out, this is probably an oversimplification of the two views. Even if a levels-of-processing view is correct, the organization of the memory storage substrates (i.e., buffers) associated with particular levels of processing is still an important question (also cf. Craik & Levy, 1976).

There are at least two points in processing where a buffer memory is thought to exist. First, stimuli are represented in a sensory buffer from which information can be extracted until the sensory trace has decayed

(Cowan, 1984, 1987; Crowder, 1982; Massaro, 1972; Sperling, 1960). Second, recognized units reside in a short-term-memory buffer (e.g., Miller, 1956) that can be accessed in memory tasks or can be used within a "working memory" system (Baddeley, 1981, 1983) to support various problem-solving activities. Baddeley also proposed that there is a "phonological memory buffer" whose contents can be mentally rehearsed, forming an "articulatory loop." This phonological buffer could be part of the short-term-memory buffer, or it could be a distinct mechanism (see Schweickert & Boruff, 1986; Zhang & Simon, 1985). However, research by Cheng (1974) at least demonstrates that there are distinct auditory and articulatory buffers with different properties.

The type of buffer relevant to the present work is a prespeech buffer that holds sets of speech items from which utterances are selected and produced. This buffer might be the same as the phonological buffer component of Baddeley's "articulatory loop." The items presumably can originate in stimulus presentations, or they can emerge solely from the subject's thought processes. The prespeech buffer concept is important to explain speech errors (Dell, 1986, p. 285). Speakers sometimes transpose phonological or lexical elements, produce planned units of speech prematurely, or produce incorrect phoneme sequences that appear to be "blends" formed from two words (Dell, 1986; Fromkin, 1973; Motley, Camden, & Baars, 1982). These errors presumably can occur when several similar items are present in the buffer concurrently and an incorrect selection from the buffer is made.

Several general assumptions about prespeech buffer storage appear to be warranted. First, a word or sound cannot be uttered until it has entered the buffer. Second, there are at least some cases in which an unwanted item enters the buffer automatically, regardless of the subject's wishes (e.g., in the Stroop task). Third, there is a selection mechanism following the buffer, which allows the subject to trace the origin of each item and decide which item is relevant to the task. Thus, a speaker may reject some of the words that occur to him and pronounce others. Items that are not selected leave the buffer (although it

is not clear whether they simply decay, are replaced, or are actively removed from the buffer). The selection mechanism sometimes errs, and the speed of selection can vary with the kind or amount of interfering input to the buffer.

Several alternative sets of characteristics for the prespeech buffer theoretically are possible, depending upon how much the buffer can hold at one time and how items are selected from the buffer. The following three models of the buffer yield alternative sets of predictions when there is both relevant and irrelevant input to the buffer concurrently.

1. The buffer might be limited to one item at a time. An item arriving when the buffer was occupied presumably would be lost; for that item to be spoken, it would have to be reactivated when the buffer was free. Interference effects would occur when an irrelevant item arrived just before the relevant item.

2. On the other hand, the buffer might be able to hold multiple items, but the mechanism that evaluates the items might process words serially. There would presumably be an ordered queue of items based on their time of arrival into the buffer. Interference effects would occur when irrelevant items preceded the target item in the queue.

3. Finally, the buffer might be able to hold multiple items at the same time, with the correct item selected through a parallel search. Interference presumably would occur because the selection process would be slowed when multiple, similar words were present. Analogous types of capacity-limited parallel search have been described by Rumelhart (1970) and Fisher (1982). Later, we will suggest that the present empirical results are most consistent with this last notion of the prespeech buffer.

Selective Attention

Selective attention can be viewed as a mechanism that permits a subset of the information present in a buffer to be processed further. However, there is currently a debate about the level of processing at which selective attention occurs (e.g., after an unanalyzed sensory buffer vs. after a buffer containing analyzed percepts). The present work is relevant to that issue.

In the early work on selective attention, subjects were to monitor or attend to a particular physical channel of input, such as one ear in dichotic listening (Broadbent, 1958). The general finding was that subjects could selectively attend to particular physical features of the stimuli, but that it was much more difficult to attend to stimuli with a particular semantic or conceptual description. Also, subjects could recall only the basic physical characteristics of the unattended material. This led to a description of selective attention as an ''early filter'' that prevented unattended stimuli from advancing in processing beyond a sensory memory buffer. (This description does not deny that selection among attended stimuli at later points in processing also can occur.)

However, subsequent research suggested that unattended information was sometimes analyzed further. For example, Moray (1959) found that subjects detected their own names in the unattended ear in dichotic listening, and Treisman (1960) found that subjects attempting to shadow (i.e., repeat) input from one ear occasionally followed a sentence inadvertently when it switched from the relevant to the irrelevant channel. To account for such results, Treisman (1960, 1964) proposed that irrelevant channels of input are only ''attenuated'' rather than filtered out. She suggested that attenuated input receives some perceptual analysis, and that this analysis is sufficient to recruit attention when the input matches a lexical unit that is activated or ''primed'' by its special significance to the subject or the current context (also see Johnston & Dark, 1986).

Other theorists (e.g., Deutsch & Deutsch, 1963; Norman, 1968) departed further from an early filter model, proposing that perceptual analysis is completed automatically and does not depend upon selective attention at all. According to these ''late filter'' theories, selective attention occurs only at the decision end of processing and involves the selection of some stimuli for action, conscious thought, or enhanced retention. The primary evidence underlying this position is that semantic attributes of unattended stimuli that do not reach awareness still seem to influence the processing of attended stimuli, in dichotic listening (e.g., Lewis, 1970; MacKay, 1973) and in visual masking (e.g., Balota, 1983; Marcel, 1983). Moreover, words paired with shock and then presented without shock in the unattended channel in dichotic listening have been found to elicit continued physiological responding, which also generalizes to words with similar sound or meaning (Von Wright, Anderson, & Stenman, 1975). However, Holender (1986) criticized all of this research in an extensive review, suggesting that there is insufficient evidence that subjects remain unaware of the allegedly unattended information.

Additional studies support the intermediate (i.e., moderate) position that unattended information receives some perceptual analysis, but less than attended information receives. Eich (1984) used a selective listening task in which some of the unattended items were disambiguated homophones (e.g., taxi-FARE). In a subsequent test session, subjects could not discriminate items that had versus had not been presented in the unattended channel. Nevertheless, when asked to spell the ambiguous items they more often used the presented versions. This perceptual analysis of unattended items was partial rather than complete, though, because spelling scores were much higher in another condition in which subjects attended to the list of homophones. Kahneman (1975) reported that there was a negative correlation between the retention of information presented simultaneously to the left and right ears in dichotic listening when the task was an effortful one (e.g., recall) but not when the task required little effort (e.g., recognition). Last, Näätänen (1986) has identified

a component of the event-related cortical potential that is very similar for attended and unattended stimuli in dichotic listening and is sensitive to detailed physical (but not necessarily semantic) properties of the stimuli. Thus, although one can fully attend to only one channel at a time, partial processing seems to occur in unattended channels.

These considerations about the nature of a prespeech buffer and selective attention lead to alternative predictions about performance in a cross-modal Stroop task, discussed below.

Predictions of the Cross-Modal Stroop Task

In the color-word interference or "Stroop" task (Stroop, 1935; see Virzi & Egeth, 1985, for a summary and theoretical account of subsequent research), a color word is presented in an inconsistent color ink and the subject is to identify the color of the ink. It is found that responding is slowed relative to a control (nonword) presentation of that color. Thus, subjects cannot ignore the irrelevant, printed information. Presumably, two color names are available in a prespeech response buffer, but the name arising from the printed word must be suppressed for correct responding to occur. In the present experiment, subjects received the same visual Stroop task, but with concurrent, irrelevant auditory color words, noncolor words, nonspeech (music) stimuli, or silence.

Predictions about selective attention. The studies suggesting that there is some processing of unattended inputs generally require that subjects distinguish between relevant and irrelevant channels of an input modality. However, some previous research (Broadbent, 1958; Treisman, 1969) indicates that attending to one of two modalities receiving stimuli is easier than attending to one of two stimulus dimensions within a modality. Therefore, it is conceivable that there would be no perceptual processing of stimuli in an irrelevant modality, such as auditory stimuli in the present cross-modal Stroop task. In that case, cross-modal interference should not occur.

Alternatively, the irrelevant input may be automatically processed enough to deposit the item in the prespeech buffer, resulting in cross-modal interference. Consistent with that suggestion, Salamé and Baddeley (1982) found that unattended speech impaired memory for written words, to an extent that depended upon the phonological similarity of the spoken and written words.

Predictions about the prespeech buffer. If there are cross-modal Stroop effects, the detailed pattern of results may help to distinguish between the three models of the prespeech buffer enumerated above. The relevant empirical questions are: (a) whether speech sounds unrelated to colors also interfere with performance, and (b) whether auditory and visual Stroop interference effects are additive (i.e., greater with interference from two modalities than with interference from either modality alone).

The amount of interference to be expected from words unrelated to colors would depend upon the manner in which items are selected from the speech buffer. One way in which there could be substantial interference from non-

color words would be if the mechanism for selecting items from the buffer were limited to an examination of one item at a time. This could occur if the buffer had only a 1-item capacity (Buffer Model 1), or if it had a multiple-item capacity but processed items serially (Buffer Model 2). In either case, the selection mechanism would have to examine each item in the buffer individually, which should take time regardless of the similarity of the interfering word to color words.

In contrast, if the mechanism for selecting a response from words in the buffer were able to process multiple items in parallel (Buffer Model 3), little or no interference would be expected from words that did not resemble color words. Prior research suggests that the ease of selecting the correct word depends upon the dissimilarity between the words in the buffer concurrently (Klein, 1964), and it should not be difficult to select color words while rejecting words that are very dissimilar.

The additivity of visual and auditory color-word interference may depend upon the capacity limits of the buffer. Nonadditivity would be expected if the buffer had a 1-item capacity limit (Buffer Model 1). According to this model, interference would occur whenever an irrelevant item was present in the buffer when the correct item arrived, because of the time and effort needed to remove the irrelevant item from the buffer before the correct item could be entered. Presumably, the subject would repeatedly engage in processing to enter the correct item until the buffer was available (analogous to a caller attempting to reach a busy phone line). However, a second interfering word arriving when the buffer was occupied by a first interfering item would be lost, causing no additional interference.

If the buffer were able to contain more than one item at a time (as in Buffer Models 2 and 3), one would expect the auditory and visual interference effects to be additive. In Buffer Model 2, additive interference would occur because it would be possible for two interfering words to be ahead of the correct word in the queue that was processed by the selection mechanism. That mechanism presumably would take time to reject each of the erroneous words, and in addition, occasionally one of these erroneous words would be mistaken for a correct response. In Buffer Model 3, on the other hand, additive interference would occur because a parallel selection mechanism has to choose among three color-word possibilities, which presumably would be more time-consuming and susceptible to error than choosing between only two possible responses (i.e., during simple Stroop interference).

Precedents to the Present Research

Previous studies (Green & Barber, 1981; McClain, 1983) have documented auditory analogues to the Stroop effect in which both the relevant and irrelevant stimulus traits were auditory, but these studies did not examine cross-modal Stroop interference. On the other hand, Thackray and Jones (1971) used a cross-modal procedure comparable to the present study, except that the response was to press a key marked with the appropriate color

name. No cross-modal effects were obtained. Houston and Jones (1967), using a mixture of nonspeech and speech sounds, actually obtained a slight release from visual Stroop effects, but separate means for specific types of sounds were not reported.

Morton (1969) did find Stroop-like interference from spoken digits in a task in which subjects were to count visually presented items. There are two problems with this evidence, however. First, the experiment was described only briefly and means were not presented. Second, the auditory conditions apparently were not counterbalanced. Although the auditory control (nondigit) conditions were presented first so that practice effects could not invalidate the finding of auditory interference, task fatigue or proactive interference across trials would invalidate this finding. For these reasons, and because counting may not operate in the same manner as color naming, the present study of cross-modal Stroop interference was conducted.

The Present Task

On each trial, the subject attempted to name as rapidly as possible a sequence of colors presented either in the form of color words or as strings of xs. During the task, the subject heard one of five presentations over earphones: (1) a random series of spoken color words, (2) repetitions of the word "the," (3) repetitions of the alphabet, (4) part of a sonata, or (5) silence. The noncolor speech materials were selected with the intent of avoiding uncontrolled associations to the color system. Klein (1964) has shown that there is Stroop-like interference from visually presented words that evoke color associations without naming the colors (e.g., "lemon, grass, fire, sky"). Other nouns or adjectives also would have some color associations, but these would be uncontrolled and might differ among subjects. In contrast, the word "the" or letters of the alphabet should have few, if any, color associations. Considerable evidence (e.g., Baddeley, 1983; Cowan, Braine, & Leavitt, 1985; Dell, 1986; Drewnowski, 1980; Salamé and Baddeley, 1982) suggests that auditory-verbal input should result in phonological sequences in a prespeech buffer regardless of the semantic or grammatical nature of that input.

The purpose of including two different non-color-word conditions was to ensure that the distinction between color and noncolor speech materials was not a simple acoustic or phonetic one. Repetitions of "the" are highly redundant, whereas repetitions of the alphabet provide more phonological variety. The music condition was included in order to determine if any effects that occur for the irrelevant speech conditions are speech-specific.

METHOD

Subjects

The subjects were 32 college students (23 women and 9 men) who received credit in an introductory psychology class. An additional subject was excluded because she could not distinguish all of the colors.

Materials

On each trial, the subject read from a page on which 100 items were printed. The page was 35.56 cm long × 21.59 cm wide, with four columns of words and lettering in lower case, about 0.7 cm high. The items were printed in red, blue, green, black, or orange ink. In one condition, the items were conflicting printed color words selected from the same five colors; in another condition, the printed items were strings of the letter x, matched in length and color of ink to the color words. Each color (and in the Stroop condition, each color word) appeared equally often on a page. The orders were randomized with the constraint that the same color could appear no more than twice in a row and, in the Stroop condition, that no color word could appear more than twice in a row. Five color-word pages and five control pages were constructed, to be used in combination with different auditory conditions.

Subjects listened to auditory stimuli via a CTR-70 Realistic tape recorder and Mura headphones (foam type). All of the speech stimuli were produced in the same male voice. On one audiotape, the five *colors* were spoken in a random order that did not correspond to any of the printed materials. Color words were spoken at a mean rate of 1.66 words/sec. Two other tapes contained noncolor speech items. On one of these, the English *alphabet* was repeated over and over at an even pace (e.g., without speeding up the sequence "l-m-n-o" as is often done). One repetition of the alphabet took 7.28 sec, for a mean rate of 3.57 letters/sec. This tape was constructed by repeating a tape loop of one alphabetic repetition. The other non-color-word tape contained the word *the* spoken at a rate of 1.25 words/sec. A single token of the word was digitized using a Zenith-110 microcomputer and an I/O Technology analog-to-digital-to-analog control board; the tape was constructed by playing this sound repeatedly. Finally, a fourth tape contained *music*, specifically, "Sonata No. 2 for Violin and Piano" by Bartok. All four of these tapes were adjusted to comfortable listening levels that the authors agreed were subjectively equivalent. In a final condition, the subject received only *silence*, but still wore the headphones.

Procedure

Testing was conducted in a sound-attenuated chamber, with the subject facing one wall and the experimenter behind him or her. The subject was instructed to ignore the sounds and the color words, and to concentrate on the color-naming task. He or she was to name the colors aloud as quickly as possible without making errors. Following a brief practice session, each subject received 10 pages of stimuli (1 page for each combination of visual and auditory conditions) with a brief rest period following each page. The subject received both trials with a particular audiotape (i.e., one trial with printed color words and one with strings of the letter x) successively. The five audiotapes were used in an order that corresponded to one row of a Latin square for each subject; the subjects received the visual Stroop and control conditions in an alternating order, beginning with color words for half of the subjects and xs for the other half. The experimenter used a stopwatch to record the time from the beginning of each trial (initiated by removing a cover from the stimulus list) to the time that the subject finished responding to the last item on the page. During the trial, the experimenter recorded errors on a listing of the correct responses.

RESULTS

Response Times

The amounts of time it took subjects to read each 100-word list were analyzed in a 2×5 analysis of variance (ANOVA) of the response times, with visual condition (color words or xs) and auditory condition (color words,

the alphabet, "the," music, and silence) as within-subject factors. The mean response times are shown in Figure 1. There was a large main effect of the visual condition caused by slower performance in the color-word condition than in the control condition [$F(1,29) = 353.66$, $p < .001$, $MSe = 101.40$]. There was also an effect of the auditory condition, apparently caused by slower performance in the color-word condition than in the other auditory conditions [$F(4,116) = 9.88$, $p < .001$, $MSe = 31.78$]. However, the interaction effect did not approach significance [$F(4,116) = 0.77$].

Post hoc, Tukey tests between pairs of means for the five auditory conditions resulted in significant differences between the color-word condition and three of the other conditions (alphabet, $p < .05$; music, $p < .01$; and silence, $p < .01$). None of the other comparisons between pairs of means were significant.

Errors

Overall, the subjects pronounced an interfering word rather than the correct color only occasionally in visual Stroop conditions (1.67% of all items) and even more rarely in visual control conditions (0.53%). Despite these low error rates, there were important differences between auditory conditions. In an ANOVA with the same factors used in the analysis of response times, there was a significant main effect of the visual condition [$F(1,29) = 35.41, p < .001, MSe = 2.75$] and of the auditory condition [$F(4,116) = 4.14, p < .004, MSe = 1.64$]. Unlike the response time analysis, in the error data there was also a significant interaction of visual and auditory conditions [$F(4,116) = 4.39, p < .003, MSe = 0.98$]. Because the standard deviations for the various conditions
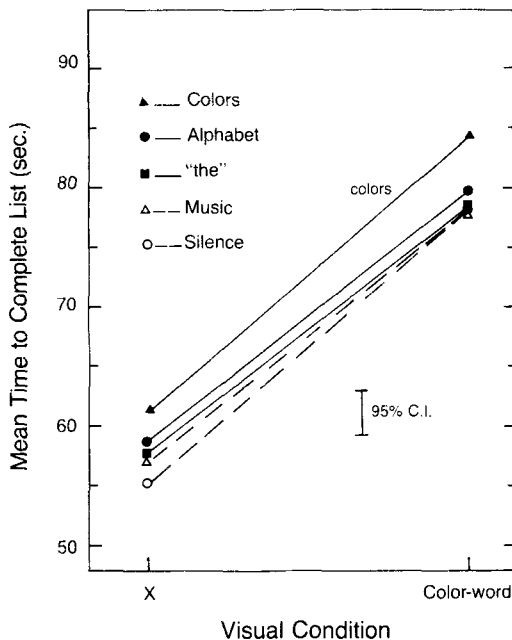
Figure 1. Mean time to name colors on a 100-item list in each condition of the experiment. The graph parameter is the auditory condition.
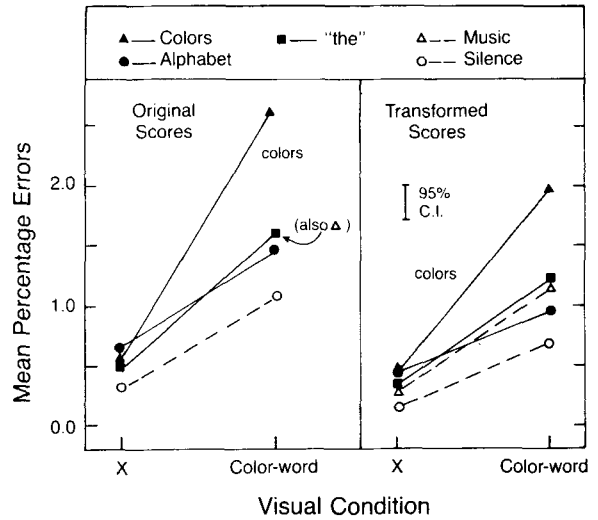
Figure 2. Mean percentage of errors on each 100-item list in each condition of the experiment. Left panel: Original percentage scores. Right panel: Estimates obtained by detransforming means obtained after a logarithmic transformation of the data. The graph parameter is the auditory condition.

were roughly proportional to the means, another ANOVA was conducted after the data were transformed according to the equation $y' = \log(y+1)$, as suggested by Myers (1972, p. 77). In this analysis, the same effects were obtained as in the previous analysis ($p < .001$ for visual condition, $p < .007$ for auditory condition, and $p < .02$ for the interaction).

The pattern of means responsible for these effects is shown in Figure 2, with the original means on the left and scores obtained by detransforming the means from the data transformation on the right. The results suggest that there were more errors in the visual Stroop condition than in the visual control condition, but also that especially many errors were made when both visual and auditory color word were present at the same time. These statements were supported by Tukey tests on the means of the transformed scores for each auditory condition, carried out separately for the visual Stroop and visual control presentations. With a visual Stroop presentation, the auditory color-word condition was found to produce significantly more errors than each of the other conditions (alphabet, $p < .01$; "the," $p < .05$; music, $p < .05$; and silence, $p < .01$). However, none of these other conditions differed significantly from one another. Moreover, with a visual control presentation, none of the auditory conditions differed significantly from one another.

DISCUSSION

In the present study, subjects performed a visual Stroop task that involved a spoken response and various types of auditory presentation. Visual Stroop interference was obtained, as one would expect, but additional effects of the auditory presentation are of primary interest. The basic findings were that (1) spoken color words interfered sub-

stantially with performance, (2) there was neither interference nor facilitation from spoken non-color-words or music, and (3) interference from spoken and written color words was additive (i.e., performance was poorer with both types of interference present than with either type alone). These findings place constraints on the characteristics of selective attention and buffer storage that will be discussed in turn.

Selective Attention in
the Cross-Modal Stroop Task

Debates about selective attention have focused on whether information is blocked from further processing before or after some (or all) perceptual analysis has been performed (Holender, 1986; Johnston & Dark, 1986; Kahneman & Treisman, 1984). Although there is evidence that at least some perceptual analysis of unattended input takes place (e.g., Eich, 1984; Näätänen, 1986), we raised the possibility that this perceptual analysis might be blocked in situations in which subjects could attend to one modality and ignore another. If so, cross-modal Stroop effects should not be obtained. The presence of these effects indicates that the irrelevant, spoken color words were processed to some degree automatically, ruling out an extreme "early filter" model of selective attention. Notice that a cross-modal Stroop effect was obtained even in the condition in which visual xs were combined with spoken color words. In that condition, all of the visual input was relevant and all of the auditory input was irrelevant, but this modality separation did not eliminate automatic perceptual analysis.

On the other hand, the results do not imply that perceptual processing runs to completion automatically, as in an extreme "late filter" theory of selective attention. There are two reasons why it would be premature to accept this late filter model. First, because the color words are highly primed by the test context in the Stroop task, a partial perceptual analysis of a spoken color word might be sufficient to activate the corresponding item in long-term memory and deposit it in the prespeech buffer. Treisman (1960) and Johnston and Dark (1986) suggested that primed items might be more easily activated than other items. Second, there is no evidence that the perceptual analysis of speech that is needed to cause cross-modal interference must include a semantic component. It is theoretically possible that a phonetic representation of the irrelevant speech item entering the prespeech buffer is sufficient to interfere with performance. Thus, the data rule out an extreme early filter model but do not distinguish between the remaining models. However, evidence reviewed above (e.g., Eich, 1984) suggests that a moderate stance, such as Treisman's attenuation model, is more viable than an extreme late filter model in which perceptual analysis is totally automatic.

The Use of a Prespeech Buffer

The data also support the notion that verbal items from both the visual and auditory modalities enter a common prespeech buffer from which a response must be selected. The most straightforward explanation for multimodal Stroop interference effects is that the presence of several color words in the buffer increases the difficulty of selecting the correct response. A similar logic was used to explain the finding (Salamé & Baddeley, 1982) that unattended speech can interfere with memory for phonologically similar printed items. In that research, also, speech appears to enter a buffer that receives input from visual and auditory modalities. The similarity of findings in the two types of experiment suggests that the relevant buffer in these tasks may be the same.

The assumption that the relevant buffer in the present task is a prespeech buffer (i.e., that it subserves a spoken response) accounts for a discrepancy between the present results and those of Thackray and Jones (1971). They presented auditory color-word interference in a visual Stroop task, but the response mode was to press computer keys marked with written color names, and no auditory interference was obtained. This sort of response may bypass the prespeech buffer. A study by Neill (1977) can be analyzed similarly. In his experiments, the distracting color name in one trial sometimes was the relevant color in the next trial (e.g., the word "red" in blue ink followed by "green" in red ink). With a vocal response (presumably based on the prespeech buffer), there was added interference from the previous unattended item. However, when a manual response was used (presumably circumventing the prespeech buffer), there was a slight facilitation rather than interference from the prior unattended item.

The importance of a prespeech buffer is compatible with a general model of the Stroop task (Virzi & Egeth, 1985) in which the compatibility of stimuli and responses is stressed. That model is based largely on evidence that, when a nonverbal, color-matching response is required, the ordinary Stroop effect is not obtained (see also McClain, 1983). Presumably, because the original Stroop task required translation of the relevant information from nonverbal to verbal form, there was competition from the written color, which did not require the same elaborate translation. In the terminology of Virzi and Egeth, a translation stage intervenes whenever the input system analyzers are incompatible with the decision and response mechanisms required by the task. In the present case, the buffer common to both modalities must occur in the response end of the processing system. Both auditory and visual verbal materials are compatible with a spoken response and do not need to go through the translation mechanism discussed by Virzi and Egeth, whereas the color of ink must be translated into verbal form for a response to occur. This translation process may allow irrelevant items to reach the prespeech buffer before the relevant item.

Discriminating among models of the buffer. One aspect of the results further constraining models of a prespeech buffer is that the effects of visual and auditory interference were additive. Subjects' response speeds were

slower when both visual and auditory color-word interference was present than when either alone was present. This result contradicts the description of a 1-item buffer offered above (Buffer Model 1), because the occupation of the buffer by one interfering word presumably would result in the loss of (and lack of an effect from) the other potentially interfering word. The additivity of visual and auditory effects suggests, instead, that the buffer has a multi-item capacity.

Another aspect of the data that constrain models of the buffer is that there was no interference effect from noncolor words. (The means displayed in Figure 2 suggest a possible, although nonsignificant, effect of noncolor words on error rates, but this potential effect was of the same magnitude for the music condition and could be attributed to a general attentional interference with the selection or response process.) According to either Buffer Model 1 or a model of the buffer in which multiple words enter the buffer but are evaluated in sequential order (Buffer Model 2), there should have been substantial interference effects of noncolor words entering the speech buffer.

Buffer Model 3 is more consistent with all of the present data. In this model, multiple items enter the buffer, and the selection mechanism evaluates the source of these words in a parallel fashion. Presumably, the amount of interference caused by any word in the buffer would depend upon the phonemic or semantic similarity of that word to a color word. This could occur because words differing greatly from color words are quickly rejected, freeing the selection mechanism to focus on more likely candidates. The noncolor words used in this experiment were quite dissimilar from color words, and would not be expected to cause substantial interference (see Klein, 1964). However, because the selection mechanism is presumed to have a limited capacity, two incorrect color words (one of a visual origin and one of an auditory origin) would slow the selection mechanism more than would a single incorrect color word. Also, with three color words competing for selection from the buffer, an erroneous selection would be made more often than in simple unimodal Stroop interference.

Buffer Model 3 also could be used to explain the difference between response times and errors in the present results (cf. Figures 1 and 2). Auditory color-word interference alone did not cause a substantial number of errors, although it did slow the subjects' responses. On the other hand, the presence of concurrent auditory and visual Stroop interference both caused a large increase in the percentage of errors and slowed responses. One account of these results assumes that each word in the buffer is tagged according to its source of origin and focuses on the nature of this tag and how it is used in the selection process. The tag must contain information about both the modality of the word and, if the modality was visual, whether the source was the written word or the color of ink. When only auditory interference is present, the selec-

tion process can simply examine the tag to determine whether the source was auditory or visual. This process could take time without leading to errors. When only visual interference is present, the selection process must examine instead whether the source was originally verbal or nonverbal. This process may be more susceptible to errors, as well as being time-consuming. Finally, when both auditory and visual sources of interference are present, the subject must process both types of information about the sources of origin, and this type of dual search apparently is even more time-consuming and error-prone.

A previous visual Stroop experiment (Klein, 1964) provides further support for a prespeech buffer that holds multiple items and a selection device that examines these items in parallel. Subjects in that experiment were to read each color word and then name the color of the ink for the same stimulus or, in another condition, name the color of the ink and then read the color word. Reading the color word first greatly facilitated color naming, whereas reading the color word second impeded color naming. According to the present account, both potential responses would reside in the prespeech buffer simultaneously. When the color word is read, it is removed from the buffer, simplifying the task of color-naming. In contrast, when the subject knows that the color word is to be read after color-naming, the selection process may be slowed because both responses are primed in the selection mechanism.

More work will be necessary before it can be determined whether the buffer described here is, in fact, identical to the phonological buffer component of Baddeley's (1981, 1983) "articulatory loop." Klapp, Greim, and Marshburn (1981) equated the articulatory loop with an auditory store instead, because they had found that relevant auditory input overcame the otherwise disruptive effect of irrelevant, silent articulation on immediate recall. However, this view does not take into account Baddeley's two-component description of the articulatory loop, which states that a phonological buffer serves as the storage component for a set of active encoding and rehearsal processes. Auditory stimuli enter the phonological buffer automatically, but visual stimuli can enter the buffer only with the assistance of an encoding process, which is disrupted by irrelevant articulation (Murray, 1968; Peterson & Johnson, 1971). In the experiments of Klapp et al. (1981), the relevant auditory input would enter the phonological buffer to be used in place of the blocked visual input.

Another unresolved issue is whether the prespeech buffer is limited in capacity, and if so, how. One must consider possible limits in both the duration of speech that can be stored and the number of items that can be stored concurrently (Schweickert & Boruff, 1986; Zhang & Simon, 1985). Within the Stroop task, it is now important to determine limits in the number of concurrent or consecutive irrelevant items for which color-naming interference summates.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

BADDELEY, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition, 10,* 17-23.

BADDELEY, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London,* B 302, 311-324.

BALOTA, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning & Verbal Behavior,* 22, 88-104.

BROADBENT, D. E. (1958). *Perception and communication.* New York: Pergamon Press.

CHENG, C. (1974). Different roles of acoustic and articulatory information in short-term memory. *Journal of Experimental Psychology,* 103, 614-618.

COWAN, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341-370.

COWAN, N. (1987). Auditory memory: Procedures to examine two phases. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds.* Hillsdale, NJ: Erlbaum.

COWAN, N., BRAINE, M. D. S., & LEAVITT, L. A. (1985). The phonological and metaphonological representation of speech: Evidence from fluent backward talkers. *Journal of Memory & Language,* 24, 679-698.

CRAIK, F. I. M., & LEVY, B. A. (1976). The concept of primary memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes. Vol. 4: Attention and memory* (pp. 133-175). New York: Wiley.

CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior,* 11, 671-684.

CROWDER, R. G. (1982). Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 8, 153-162.

DELL, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review,* 93, 283-321.

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

DREWNOWSKI, A. (1980). Attributes and priorities in short-term recall: A new model of memory span. *Journal of Experimental Psychology: General,* 109, 208-250.

EICH, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition,* 12, 105-111.

FISHER, D. (1982). Limited-channel models of automatic detection: Capacity and scanning in visual search. *Psychological Review,* 89, 662-692.

FROMKIN, V. A. (Ed.) (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

GREEN, E. J., & BARBER, P. J. (1981). An auditory Stroop effect with judgments of speaker gender. *Perception & Psychophysics,* 30, 459-466.

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

HOUSTON, B. K., & JONES, T. M. (1967). Distraction and Stroop color-word performance. *Journal of Experimental Psychology,* 74, 54-56.

JOHNSTON, W. A., & DARK, V. J. (1986). Selective attention. In M. R. Rozenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 37, pp. 43-76). Palo Alto, CA: Annual Reviews.

KAHNEMAN, D. (1975). Effort, recognition and recall in auditory attention. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 65-80). New York: Academic Press.

KAHNEMAN, D., & TREISMAN, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention.* New York: Academic Press.

KLAPP, S. T., GREIM, D. M., & MARSHBURN, E. A. (1981). Buffer storage of programmed articulation and articulatory loop: Two names

for the same mechanism or two distinct components of short-term memory? In J. Long and A. Baddeley (Eds.), *Attention and performance IX.* Hillsdale, NJ: Erlbaum.

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

LEWIS, J. L. (1970). Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology,* 85, 225-228.

MACKAY, D. G. (1973). Aspects of a theory of comprehension, memory, and attention. *Quarterly Journal of Experimental Psychology,* 25, 22-40.

MARCEL, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology,* 15, 197-237.

MASSARO, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review,* 79, 124-145.

MCCLAIN, L. (1983). Stimulus-response compatibility affects auditory Stroop interference. *Perception & Psychophysics,* 33, 266-270.

MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

MONSELL, S. (1984). Components of working memory underlying verbal skills: A "distributed capacities" view. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes.* Hillsdale, NJ: Erlbaum.

MORAY, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology,* 11, 56-60.

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

MOTLEY, M. T., CAMDEN, C. T., & BAARS, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior,* 21, 578-594.

MURRAY, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology,* 78, 679-684.

MYERS, J. L. (1972). *Fundamentals of experimental design* (2nd ed.). Boston: Allyn & Bacon.

NÄÄTÄNEN, R. (1986). Processing of the unattended message during selective dichotic listening. *Behavioral & Brain Sciences,* 9, 43-44.

NEILL, W. T. (1977). Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception & Performance,* 3, 444-450.

NORMAN, D. A. (1968). Toward a theory of memory and attention. *Psychological Review,* 75, 522-536.

PETERSON, L. R., & JOHNSON, S. T. (1971). Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning & Verbal Behavior,* 10, 346-354.

RUMELHART, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology,* 7, 191-216.

SALAMÉ, P., & BADDELEY, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior,* 21, 150-164.

SCHWEICKERT, R., & BORUFF, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 12, 419-425.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychological Monographs,* 74(11, Whole No. 498).

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 18, 643-662.

THACKRAY, R. I., & JONES, K. N. (1971). Level of arousal during Stroop performance: Effects of speed stress and "distraction." *Psychonomic Science,* 23, 133-135.

TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology,* 12, 242-248.

TREISMAN, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *The American Journal of Psychology,* 77, 533-546.

TREISMAN, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.

VIRZI, R. A., & EGETH, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, 13, 304-319.

VON WRIGHT, J. M., ANDERSON, K., & STENMAN, U. (1975). Generalization of conditioned GSRs in dichotic listening. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5). New York: Academic Press.

ZHANG, G., & SIMON, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, 13, 193-201.