ITEM-SPECIFIC CONGRUENCY EFFECTS IN NONVERBAL AUDITORY STROOP

by

Launa Cheryl Leboe

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology

University of Manitoba

Winnipeg, Canada

RUNNING HEAD: NONVERBAL AUDITORY STROOP

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

Item-Specific Congruency Effects in Nonverbal Auditory Stroop

by

Launa Cheryl Leboe

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of

Manitoba in partial fulfillment of the requirement of the degree

of

Master of Arts

ABSTRACT

In the current study, participants judged either the location or the frequency of a single tone that was defined by both features. The response associated with the irrelevant feature was either consistent (congruent trials) or inconsistent (incongruent trials) with the required response. The results of the first experiment showed that participants were slower to judge the relevant feature on incongruent trials than on congruent trials. This finding suggests that an auditory Stroop effect for nonverbal sounds results from the inability to ignore an irrelevant acoustic feature. In the second experiment, the likelihood that the response associated with the irrelevant dimension was congruent with the required response was manipulated. Participants were slower to respond on incongruent trials than on congruent trials when the response associated with the irrelevant feature was likely to be consistent with the required response, but were slower to respond on congruent trials than on incongruent trials when the response associated with the irrelevant feature was likely to be inconsistent with the required response. This item-specific congruency effect suggests that the extent to which the response information associated with an irrelevant acoustic feature influences performance may be flexibly controlled in accordance with the likelihood that it will be diagnostic of a correct response.

Acknowledgements

I thank Dr. Todd Mondor for his guidance and mentorship during my time as a Master's student, as well as for his feedback and assistance in the preparation of this manuscript. In addition, I thank the other members of my supervisory committee, Dr. James Hare and Dr. Murray Singer, for their helpful comments at the various stages of this Thesis. I also need to thank my husband, Jason, and our daughter, Delica, for their unwavering support throughout this process.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

ITEM-SPECIFIC CONGRUENCY EFFECTS IN NONVERBAL AUDITORY STROOP

OVERVIEW OF PROJECT

The Stroop effect provides a model case for examining whether the response associated with an irrelevant feature will interfere with performance of a primary task. Although there has been a long history of studying interference effects in the auditory domain (e.g., Broadbent, 1954; Cherry, 1953, Hall & Blasko, 2005), there have been relatively few demonstrations of an auditory Stroop effect defined as a direct conflict between two acoustic dimensions. Previous demonstrations of an auditory Stroop effect have been founded on interference between the meaning of a spoken word and an acoustic feature. For example, Hamers and Lambert (1972) presented the spoken words 'high' and 'low' in either a high or low pitch voice, and the participants' task was to categorize words according to their pitch (high vs. low). Similarly, Green & Barber (1981, 1983) presented the words 'man' and 'girl' spoken in either a male or female voice and participants categorized the words according to the gender of the speaker (labeled man and girl). In both studies, participants judged the relevant acoustic feature (pitch; gender of speaker) more slowly when it was inconsistent with the meaning of the word than when it was consistent with the meaning of the word. These and similar studies (Green & Barber, 1981, 1983; Hamers & Lambert, 1972; McClain, 1983; Morgan & Brandt, 1989) represent demonstrations of an auditory Stroop effect arising from an inability to ignore the semantic content of a spoken word when pitch classification is required. In this way, the auditory Stroop effect appears to be quite similar to the classic visual Stroop effect (e.g., Green & Barber, 1981, 1983; Stroop, 1935). For example, in his classic demonstration, Stroop (1935) showed that participants are particularly slow to identify the colour of ink in which incongruent colour words are printed (e.g., the word RED printed in green ink).

The current study was designed to explore the possibility that auditory Stroop interference may occur for two nonverbal acoustic features. Previous research provides some reason for believing that such nonverbal Stroop interference effect may occur. Specifically, Mondor, Zatorre, & Terrio (1998) reported that the time required to classify a single sound by location (central or peripheral) was lengthened when pitch varied unpredictably and that pitch classifications (high or low) were accomplished more slowly when location varied unpredictably.

The results reported by Mondor et al. (1998) suggest that listeners are unable to attend exclusively to either location or to pitch. If this interpretation is correct, then it may be that the responses associated with location and pitch will interfere as well. This is an important issue because most theoretical accounts of Stroop interference are based on the notion that automatic processing of word meaning interferes with colour identification (e.g., LaBerge & Samuels, 1974; Logan, 1988; MacLeod & Dunbar, 1988; Neely & Kahan, 2001; but see Besner & Stolz, 1999; Danziger, Estevez & Mari-Beffa, 2002; Kahneman & Henik, 1981). Obviously, evidence that Stroop interference may arise from a conflict between two nonverbal dimensions would require quite a different theoretical account.

In the experiments described below, participants were presented with high- and low-pitched sounds either from an upper or a lower speaker and were required to categorize sounds according to their location or pitch. In such a situation, evidence of Stroop interference would be apparent if participants experience greater difficulty in judging sounds when the response associated with the irrelevant feature is inconsistent with the required response than when it is consistent with the required response. The first experiment did yield such evidence and the second experiment was designed to explore the possibility that the degree to which the influence of the response associated with an irrelevant feature may be flexibly controlled.

The above section provides a brief overview of the content of this thesis. In the next section, I provide historical background illustrating the significance of interference effects for the field of cognitive psychology. Specifically, under the heading, *Processing Limitations and Selective Attention*, I discuss the way in which interference effects may provide insights regarding both limitations of human cognitive processing and mechanisms of selective attention. Next, under the heading, *Theoretical Accounts of the Stroop Effect*, I provide a summary of the ground-breaking work of J. R. Stroop (1935) who first reported that participants are slow to name an ink colour when it is paired with a incongruous colour word, and discuss the various theoretical accounts that have been suggested to explain this phenomenon. Finally, under the headings, *The Auditory Stroop Effect* and *A Nonverbal Auditory Stroop Effect*, I discuss modifications of Stroop's paradigm that have been used to investigate interference effects in the auditory domain, and the rationale for hypothesizing that a nonverbal version of this effect may occur. As well, I present two experiments, that in combination, shed new light on these matters.

Processing Limitations and Selective Attention

Our cognitive system places strict limits on the ability to process information. Usually, the amount of sensory input available at any given moment considerably exceeds the amount of information that can be processed in detail. This processing limitation formed a central component of some of the earliest investigations of cognitive psychologists. The practical importance of this issue motivated the first studies of selective attention. During World War II, there was keen interest in understanding the information processing abilities of radar operators and of airplane pilots who were required to monitor multiple stimulus inputs. Broadbent (1954, 1958) used a dichotic listening task to simulate processing of information from multiple sources to try to understand this problem.

Broadbent (1958) presented participants with a sheet of paper divided into numbered sections. Within the numbered sections, different shapes (a circle or a cross) were printed. Participants were required to use this sheet to answer questions that were presented through headphones. For example, the participant could have been asked; "Is there a circle in section one?", and the participant would answer "Yes" or "No". When only one question was presented at a time, participants had near perfect performance. However, when two questions were presented simultaneously in different channels (i.e., in different ears), such as "Is there a cross in section 1?" and "Is there a circle in section 5?", response accuracy decreased significantly for both questions. This finding led Broadbent to wonder whether the difficulty in answering questions presented simultaneously was due to a sensory limitation or to a central processing limitation. Specifically, since the questions entered the ears at the same time in the dual-task condition, the inputs might have masked one another at the sensory level. Alternatively, interference might have originated at a higher level of cognitive processing. Specifically, participants had difficulty either in processing the meaning of the two questions presented simultaneously or in generating a response to both questions.

In an effort to distinguish between the influence of sensory and central mechanisms on a divided-attention task, Broadbent (1958) alternately presented words from each question, ensuring no sensory overlap between the two messages. If the difficulty in answering two questions presented simultaneously resulted from the messages entering the ears at the same time, then eliminating overlap should have allowed participants to perform quite well. However, Broadbent found that alternating the messages led to the same, poor, level of performance as when the words from both questions were presented simultaneously. These results appear to reveal a constraint on the ability of people to perform multiple tasks separate from sensory influences. At the time, this demonstration of a central limitation in human information-processing represented a most important contribution to the emerging field of cognitive psychology.

Just as Broadbent was inspired to study the demands on cognitive resources placed on radar operators, some researchers today study the attentional demands faced by air traffic controllers (Rantanen & Levinthal, 2005). Disasters have been known to occur when the cognitive demands placed on air traffic controllers exceed their capacity to process information. In one such instance in 1976, one air traffic controller in the former Yugoslavia was responsible for simultaneously monitoring 11 aircraft. Unfortunately, the attentional demands of the task exceeded the controller's capacity to deal with the information and two of these planes collided, killing 176 passengers and crew (Barber, 1988). Similar examples abound (Pape, Wiegmann, & Shappell, 2001; Pounds, Scarborough & Shappell, 2000; Shappell & Wiegmann, 1996, 2000).

Although not everyone is responsible for the safety of passengers on an airplane, information processing limitations do affect everyone. Performance of any task may be impaired if the attentional demands of that task exceed the cognitive resources that are currently available (e.g., Kahneman, 1973). For example, errors in driving are more likely to occur when some attentional resources are simultaneously allocated to another task. Researchers have confirmed this source of impaired driving in the laboratory, demonstrating that talking on a cell phone can seriously disrupt the ability to obey traffic rules (Strayer, Drews, & Johnston, 2003; Strayer & Johnston, 2001). In general, when a secondary task reduces the availability of cognitive resources, performance of a primary task is often impaired (Bookbinder & Osman, 1979; Broadbent, 1958; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Long, 1977; Rinder, 1974). Drivers who choose to engage in secondary tasks may well endanger their own lives as well as the lives of innocent bystanders.

It appears that several factors, such as sleep deprivation and drug and alcohol intoxication, may reduce an individual's capacity to process information (e.g., Lieberman, Tharion, Shukitt-Hale, Speckman, & Tulley, 2002; Pilcher & Walters, 1997; Taylor & McFatter, 2003; Thomasius et al., 2003). For instance, it appears that driving while sleep-deprived can be as dangerous as driving while performing additional tasks. For example, in the United States from 1989 to 1993, fatigue was ruled to be a contributing factor in 56,000 car crashes, 1,544 of these with fatalities, (Knipling & Wang, 1994). Drug and alcohol intoxication also seem to reduce an individual's capacity to process information (Mitchell, 1985; Moskowitz & McGlothlin, 1974). This evidence suggests that the capacity to process information is not fixed. Alertness, drug, and alcohol intoxication are just a few of the factors that seem to reduce the overall capacity of an individual to process information.

Although there is a limitation in processing information, this problem may be partially overcome by selective attention to one or more stimulus inputs to the exclusion of others. Of course, if people could completely block processing of irrelevant information, all processing resources would be allocated to the primary task, with none being consumed by other sources of input. Nevertheless, it is easy to generate real-life examples that reveal that the ability of people to avoid processing irrelevant information is less than perfect. For example, any student who has sat in a lecture hall will know how hard it is to ignore nearby disruptive, talkative students. Even though the student knows the conversation is irrelevant, and that paying attention to it will impair their ability to follow the lecture, their attention may still be drawn to the conversation. Similarly, if people try to attend to multiple tasks at once, such as the air traffic controller who was required to monitor 11 aircraft, then attentional resources will be divided among the tasks and as a result performance will also be impaired.

Researchers have sought to identify the factors that determine whether irrelevant information will disrupt performance of a relevant task. The dichotic listening task (Broadbent, 1958; Cherry, 1953; Moray, 1959) provided one of the earliest methods for identifying the features that make a relevant message difficult to attend to, and an irrelevant message difficult to ignore. In this task, participants typically wear headphones with different messages presented to each ear. In the selective attention version of this task, the participant repeats aloud, or "shadows," one message, while ignoring the other message. Research has demonstrated that the selective attention task becomes more difficult when the unattended message is physically similar to the attended message, either in spatial location (Cherry, 1953) or acoustic similarity based on bandwidth filtering (Spieth, Curtis, & Webster, 1954). Shadowing efficiency has also been shown to decrease as the semantic similarity of the attended and unattended messages increases (Poulton, 1953; Webster & Thompson, 1954).

Cherry (1953) demonstrated that participants could not completely ignore information in an unattended channel. In his study, participants performed a dichotic listening task in which they shadowed one message while ignoring another. At some point during each trial, the message in the unattended ear could either switch language, speaker gender, or become reverse speech. Although participants were not able to report the semantic content of the message from the unattended ear, they were able to report when the gender of the voice changed; and when questioned, many also noticed something 'odd' when the message changed to reverse speech. If participants had entirely blocked out the irrelevant information, then detecting even a change in voice should have been impossible. Further, research by Moray (1959) revealed that participants were frequently able to hear their own name in the unattended message, suggesting that the salience of the information in the unattended ear can be an important determinant of whether or not it will be detected. Wood and Cowan (1995) replicated Moray's result, but found that participants only heard their name in the ignored message about 35% of the time. Thus, even when information in the unattended channel is as salient as one's own name, it will not always be detected. In any event, this evidence demonstrates that people can not always completely ignore irrelevant information.

In a further demonstration of the inability to completely block processing of irrelevant information, Treisman (1960) had participants shadow a meaningful message in one ear, but then switched the message to the unattended ear part way through the trial (the meaningful message is indicated by italics). For example, a participant might be presented with "*While we were talking she would come and go with rapid glances at us leaving on her passage an impression of grace and* is idiotic idea of almost there is cabbage a horse" in the to-be-shadowed ear, and "The camera shop and boyhood friend from fish screamed loudly singing men and then it was jumping in the tree *charm and a distinct suggestion of watchfulness.*" in the to-be-ignored ear. Participants sometimes shadowed the semantically connected text even though this meant that their attention switched from one channel to the other. Surprisingly, participants who did shadow the semantically connected text were often not aware that they had switched the ear they were shadowing. This finding shows that information in the unattended channel was processed to the level of semantics and that some people have a tendency to use verbal meaning to guide the focus of auditory attention. Whereas such a bias may facilitate coherent processing of conversations, it may also interfere with attempts to treat a source of information as irrelevant on the basis of a physical difference alone. Thus, an apparent difficulty in ignoring semantic content impairs the ability to perform the primary task of shadowing the attended message.

The Stroop effect is a well-known phenomenon in visual information processing that also provides evidence that it can be difficult to ignore semantic content. Participants in Stroop's (1935) original study were required to name the colour of ink in which words were printed. In some conditions, colour words were presented in an ink colour with which they were inconsistent. For example, participants could be presented with the word "red" printed in green ink with the requirement to respond "green". Stroop conducted three experiments in his classic study. The stimuli consisted of four cards with 100 words printed on each card. The two experimental cards were comprised of colour words printed in incongruent ink colours. The two control cards were comprised either of colour words printed in black ink or of coloured rectangles. Stroop measured the total amount of time required to read a list of words or to name the ink colour of a list of items. Interference was measured as the difference in reading or naming times between the experimental and control cards.

In the first of Stroop's experiments, participants were to read the colour words aloud, while ignoring the incongruent ink colour. In this case, the control cards included only colour words presented in black ink, thereby providing the baseline for reading colour words in the absence of any inconsistent colour information. Stroop found no significant difference in the time required to read words in the experimental and control conditions. Therefore, he concluded that colour processing does not interfere with word reading. In the second experiment, participants were required to name the ink colour, while ignoring an incongruent colour word. For this experiment, the control cards included only coloured rectangles, thereby providing a baseline for colour naming in the absence of any inconsistent word information. Stroop found a significant amount of interference for naming the ink colour when it was presented in the context of incongruent colour words. This effect of irrelevant word meaning on naming the ink colour of words is appropriately now known as the Stroop effect.

What would cause the asymmetric pattern of interference that Stroop (1935) observed? If interference resulted simply from the presence of a stimulus dimension that conflicts with the appropriate response, then there should have been equal interference for both the word reading and colour naming tasks. One explanation may be derived from Cattell's (1886) demonstration that words are read faster than colours are named. According to Cattell (1886, p. 65), "this is because, in the case of words and letters, the association between the idea and name has taken place so often that the process has become automatic, whereas in the case of colors and pictures we must by a voluntary effort choose the name" (see Posner & Snyder, 1975; Schneider & Shiffrin, 1977 for similar views on an automatic/voluntary distinction). From this point of view, the Stroop effect occurs because of a difference in the amount of practice people have with word reading versus colour naming. Stroop's asymmetric pattern of interference would be expected because the process of word reading brings to mind a colour label faster than does the process of colour naming. As a result, the need to resolve interference between two conflicting responses would only occur when the task is colour naming. When the task is word reading, the correct response could be generated and produced before interference from an inconsistent ink colour occurred.

To test this idea, Stroop (1935) conducted a third experiment in which participants practiced naming colours for eight days. He reasoned that if the asymmetric pattern of results apparent in the previous experiments was due to more practice with word reading than with colour naming, then extended practice with colour naming should have reduced the amount of interference by the end of the eight day training session. Again, interference was measured by the difference in time to name the ink colour of incongruent words or of rectangles. Over eight days of training, interference did indeed decrease from 49.6 to 32.8 seconds. Therefore, while practice with colour naming reduced the degree of interference, the inconsistent colour word still had a substantial effect on colour naming performance. However, it is important to note that participants only practiced colour naming for eight days, whereas they had a lifetime of word reading experience. It is possible, then, that additional practice may have equated the speed of word reading and of colour naming, and the Stroop effect may have been eliminated.

Interestingly, not only did Stroop demonstrate the effect of practice on colour naming, but he also demonstrated a reverse Stroop effect in which ink colour interfered with word reading. Specifically, after eight days of practicing colour-naming, participants were slower to read words printed in inconsistent ink colours than they were on day one. Taken together, Stroop interpreted his results as supporting a differential practice account. That is, reading is much more practiced than naming colours and this difference in practice is responsible for the asymmetric pattern of interference. However, Stroop's account is only one of several competing explanations of the effect he discovered. The next section provides a detailed discussion of the most dominant accounts of the Stroop effect.

Theoretical Accounts of the Stroop Effect

When originally published, Stroop's experiments did not have a great impact. It was not until Broadbent and other researchers began investigating selective attention that the importance of Stroop's experiments came to be recognized. The broader significance of Stroop's research arises from the clear illustration his studies provide that the ability of people to attend selectively is limited. The Stroop effect, therefore, provides a model demonstration of the inability of people to completely ignore irrelevant information. What makes this point most convincingly is that, in this circumstance, the failure to ignore irrelevant information impairs performance of the relevant, primary task.

Relative Speed of Processing Account

The modern theoretical account most consistent with Stroop's own interpretation of his data is the relative speed of processing approach (MacLeod, 1991). This account assumes that word reading is a faster process than colour naming, as demonstrated by Cattell (1886) and Fraisse (1969). It is further assumed that there is a limited-capacity response channel in which responses enter an output channel one at a time, with order determined by speed of entry. According to this account, given that people read words more quickly than they name colours, the response from word reading is available before the response from colour naming. The reason that interference occurs when colour naming is required is because of the time required to resolve the conflicting responses. That is, if word reading is faster than colour naming, then the response generated from word reading will generally be accessed first, and some time will be required to reject that response when it is incorrect.

One prediction based on the speed of processing account is that the Stroop effect should be reduced by any factor that slows word reading or speeds colour naming. Using this logic, Dunbar and MacLeod (1984) presented words printed either in a normal orientation or upside-down and backwards. Presenting words upside-down and backwards should slow word reading, making that process more similar in speed to colour naming. If speed of word processing is the critical factor in the Stroop effect, then normally oriented words should produce a much larger Stroop effect than upside-down and backwards words. However, contrary to this prediction, Dunbar and MacLeod (1984) found that words presented upside-down and backwards produced as large a Stroop effect as words presented in a normal orientation. This counterintuitive finding directly challenges the speed of processing account.

Automaticity Account

The automaticity account of the Stroop effect is based on the assumptions that there is an attentional resource limitation, and that cognitive processes differ in the amount of resources required. According to Schneider and Shiffrin (1977), there are two levels of cognitive processing: automatic and controlled. Automatic processing is associated with tasks that are highly practiced, such as word reading. This type of processing is effortless and can be performed unconsciously. In addition, automatic processing can occur in parallel. That is, people can process multiple inputs simultaneously. In contrast, controlled processing is associated with difficult or unfamiliar tasks such as identifying ink colour in a Stroop paradigm. This type of processing requires substantial mental effort and must be performed consciously. As a result, controlled processing is carried out in a serial manner in that only one item can be processed at a time.

In their classic study, Schneider and Shiffrin manipulated both task difficulty and set size. The participants were presented a rapid sequence of 20 frames. On each frame, there were four sections, each of which could contain a letter, a number, or a set of dots. At the beginning of each trial, participants were instructed to look for between one and four targets among a distractor set of between one and four items. At the end of the trials, participants were required to decide if the target or targets had been presented. For example, a participant could be required to look for the target letters A, B, Y, and Z. After the frames were presented, the participant would answer whether any of the four targets had been presented. The difficulty of the task was defined both by the size of the set of target letters and by the relationship between target and distractor sets. In a consistent mapping condition, members of one category (e.g., letters) were always the targets, and members of another category (e.g., numbers) were always the distractors. In a varied mapping condition, members of either of two categories (numbers or letters) could be a target or a distractor on any given trial.

Schneider and Shiffrin found that, with practice, the accuracy of detecting targets in the consistent mapping condition was independent of both the number of targets and the number of distractors. As previously mentioned, the ability to process information in parallel is one of the characteristics of automatic processing. In contrast, even with the same amount of practice, the accuracy of detecting targets in the varied mapping condition decreased as size of the target set increased. In this case, participants could not perform the search automatically (i.e., in parallel) because the targets for one trial could become distractors on the next trial. This result is characteristic of controlled processing in that each item must be processed in a serial fashion.

According to the automaticity account of the Stroop effect, our extensive practice reading words has led to this process becoming automatic. In contrast, because we are not often required to name the colour of objects, colour naming is a controlled process. Given this fundamental difference between these two processes, the asymmetric pattern of

Stroop interference is not surprising (MacLeod, 1991). In a typical Stroop display, participants cannot avoid reading a word, even though doing so may impair their performance. In contrast, when required to read a word, participants do not automatically process the colour of ink in which a word is printed.

One assumption of the automaticity explanation is that, when controlled and automatic processes lead to conflicting responses, the controlled process should not interfere with the automatic process. Automatic processes produce responses effortlessly, making them immune to conflict from a process that must be initiated deliberately under conscious control. Conversely, an automatic process will interfere with a controlled process (MacLeod, 1991). Because responses derived from an automatic process will be generated more quickly than responses generated by a controlled process, there will be opportunity for response conflict to occur. A further assumption of the automaticity account is that extensive practice can reduce the attentional resources required to perform it. If this is true, then extensive practice with colour naming should cause colour processing to interfere with word naming in the same way that word processing has been shown to interfere with colour naming in the Stroop task. As discussed above, the results obtained by Stroop (1935) when he gave participants eight days of practice in colour naming appear to support this prediction because word reading was slower after eight days of practice than it had been initially. This suggests that with extensive practice, inconsistent ink colour impaired performance on word reading.

MacLeod and Dunbar (1988) sought to test whether practice with colour naming can increase its interference with word naming. They trained participants to associate a specific colour name with each of four unique shapes. On 'congruent' trials, participants were presented with the shape printed in the same colour as the colour name with which it was associated. For example, the shape that was paired with the name 'blue' would be printed in blue ink. On 'incongruent' trials, participants were presented with a shape that was printed in a different colour than the colour name with which it was paired. For example, the shape that was paired with the name 'blue' would be printed in red ink. Participants were tested on both their speed in identifying the colour of the shapes (colour naming task) and in naming the colour they had been taught to associate with the shape (associate generation task).

Early on in the experiment, participants were particularly slow to perform the associate generation task when shape colour was incongruent. However, participants were not slow to perform the colour naming task when the colour associated with the shape was inconsistent. This is not particularly surprising given that colour naming would be a much more practiced task than the associate generation task. This finding is consistent with the automaticity account in that a process that requires relatively more attentional resources does not interfere with a process that requires less attentional resources. However, halfway through the experiment, performance on incongruent trials was found to be equally poor whether the task required colour naming or associate generation. Moreover, by the end of the experiment, whereas performance on the associate generation task was actually equivalent on incongruent and congruent trials, speed of colour naming was still impaired on incongruent trials. These findings support the second prediction of the automaticity approach; namely, that the associate generation task should interfere with colour naming as it became increasingly automatic. Presumably, with practice, the associate generation task actually became more automaticized than the colour naming task. Thus, MacLeod and Dunbar's (1988) findings suggest that interference effects are influenced by practice. Specifically, the amount and direction of interference is determined, at least in part, by the relative automaticity of the relevant and irrelevant tasks.

The automaticity account is based on the notion that Stroop interference is caused by the automaticity of word reading. Therefore, another prediction that can be drawn from this explanation is that the Stroop effect should occur when a colour word and an ink colour are presented concurrently but are spatially separated. Specifically, if words are read automatically, then the spatial integration of the ink colour and the word should not be required to produce the Stroop effect. To test this prediction, Kahneman and Henik (Experiment 3, 1981) designed an experiment in which two words, one a colour name (red, pink, blue, or green) and one a neutral word (most, cute, shoe, or long) were presented on either side of a fixation cross. One of the words appeared in white, and the other was presented in one of four colours (red, pink, blue, or green). The participants' task was to name the ink colour of the word that was not printed in white.

The critical manipulation in this experiment was whether the colour word or the neutral word was printed in coloured ink. According to the automaticity account, if word reading is strongly automatic, then the magnitude of the Stroop effect should be equivalent regardless of whether the ink colour was presented in the context of a neutral word or of a colour word because both words were presented simultaneously. Surprisingly, however, interference was only observed when the coloured ink was presented in the context of a colour word. Kahneman and Henik (1981) interpreted their findings as challenging the view that word reading is strongly automatic. Rather, they suggested that the interference observed resulted from the allocation of spatial attention to the word presented in coloured ink, and not from the automatic reading of both words. If both words had been read automatically, then there should have been the same magnitude of the Stroop effect regardless of which word was presented in coloured ink. The fact that there was no Stroop effect when the neutral word was presented in coloured ink suggests that the colour word was not read on these trials.

Besner and Stolz (1999) provided additional evidence challenging the assumption that word reading is strongly automatic. In their study, participants were required to name the colour of a single letter within colour words that otherwise were printed in white. Besner and Stolz (1999) found that the Stroop effect can be eliminated by requiring participants to name the colour of a single letter. The study reported by Kahneman and Henik (1981) suggests that attention should be drawn to the location of the distracting colour word before the Stroop effect can be observed. This condition was met in Besner and Stolz's procedure because participants were required to direct attention to the location of the word in order to name the colour of one of its letters. Thus, even when spatial attention is drawn to a colour word, there are situations in which that word will not interfere with colour naming. This finding is problematic for an explanation of the Stroop effect that emphasizes interference caused by automatic word reading.

Motivated by Besner and Stolz's (1999) research, Danziger, Estevez and Mari-Beffa (2002) investigated the effect of the location of a single coloured letter on the magnitude of Stroop interference. They measured the amount of interference that resulted from colouring an entire colour word, its first, middle, or last letter. These researchers found that the amount of Stroop interference did not depend on whether participants were required to name the colour of the entire word or of only the last letter (52 ms vs. 63 ms, respectively). However, both of these conditions produced significantly more interference than when participants were required to name the ink colour of either the first or middle letter (23 ms vs. 32 ms, respectively). If words were read completely automatically, then interference from word reading should have been equivalent in all four conditions. If word reading cannot be considered an automatic process, then it is difficult to sustain the explanation of the Stroop effect provided by the automaticity account.

Even though there are problems with both the speed of processing and the automaticity accounts, any explanation of the Stroop effect must incorporate at least some of their premises. Specifically, it appears clear that the requirement to ignore words in the Stroop task must conflict with a lifetime of experience reading words. Studies by Stroop (1935) and MacLeod and Dunbar (1988) reveal that a task that does not normally cause interference can be made to do so with extensive practice. In addition, the notion that some processes require fewer attentional resources than others is often a useful heuristic. Although it may be possible to design situations that reduce the influence of word reading, it seems clear that people read words relatively automatically despite any intentions to do otherwise. Indeed, the Stroop effect could not occur at all if people could avoid word reading whenever they wished. It also appears that people generate colour labels less automatically than they identify words. Interference from an incongruent ink colour on word reading is minimal to non-existent (MacLeod, 1991; Stroop, 1935). It appears that people can avoid processing colour identity if it is not relevant to the current task. Thus, the best approach to finding an accurate theoretical explanation of the Stroop effect might be to seek common ground between the speed of processing and automaticity accounts. Both accounts treat the source of interference as arising from differential experience with word reading and colour naming. At a more basic level, this makes the Stroop effect an example of how performance may be impaired when prior learning is inappropriate for current task demands. Thus, the Stroop effect may fall under a broader category of Transfer-Inappropriate Processing (or TIP, see Neill & Mathis, 1998; Wood & Milliken, 1998).

<u>Stroop Interference and the Proportion-Congruent Effect</u>

One model of the Stroop effect which incorporates aspects of both the speed of processing account and the automaticity account is the parallel distributed processing model (or PDP model) proposed by Cohen, Dunbar, and McClelland (1990). One of the tenets of this model is that colour naming and word reading have separate pathways that both lead to a response mechanism. Further, a mechanism within the PDP model minimizes response conflict by modulating the strength of the colour naming and word reading pathways. Specifically, when faced with interference between the responses generated by colour naming and by word naming which arises when incongruent colour words are presented, the mechanism adjusts by strengthening the pathway responsible for processing ink colour and weakening the pathway responsible for processing word identity. As a consequence, if this interference occurs often across trials, then the magnitude of Stroop interference should be reduced. Suppression of word reading will cause the identity of incongruent colour words to have relatively less influence on response generation. This prediction is consistent with the common observation that the magnitude of the Stroop effect decreases as the proportion of incongruent trials within an experimental session increases (see Lindsay & Jacoby, 1994; Logan, 1980; Logan & Zbrodoff, 1979; Lowe & Mitterer, 1982; Shor, 1975; Tzelgov, Henik, & Berger, 1992). This finding is known as the 'proportion-congruent effect' (Logan, 1980). The model proposed by Cohen et al. (1990) is based on the assumption that the proportion-congruent effect occurs because the extent to which participants allow word reading to contribute to response selection is reduced (see also West, 1999). By this view, if incongruent trials are common across the experiment, then the contribution of word reading will be reduced relative to when incongruent trials are less common.

A similar approach was adopted by Lindsay and Jacoby (1994) who proposed that the Stroop effect occurs because the independent processes of word reading and colour naming operate simultaneously to produce a response. By using a deadline procedure and a process dissociation method, these researchers attempted to quantify the separate contributions of word reading and colour processing to performance of the Stroop task. The deadline procedure requires that participants respond to an item within a specific time period (an 800 ms time limit was used in their study; responses that were not made before the deadline were not analyzed).

In order to quantify the relative contribution of word reading and colour naming to performance, Lindsay and Jacoby (1994) measured the accuracy of responding on congruent and incongruent trials, rather than the conventional method of measuring response latencies. They used the following equation to predict performance on congruent trials: p(correct|congruent) = p(Word Response) + p(Colour Response) – [p(Word Response) x p(Colour Response)]. This equation is based on the recognition that a correct response to a congruent item can be derived from either the colour naming process or the word reading process. In contrast, they used a different equation to predict performance on incongruent trials: p(correct|incongruent) = p(Colour Response) x [1-p(Word Response)]. This equation is based on the recognition that a correct response on an incongruent item will only be produced if participants rely on the colour naming process.

The validity of using these equations to compute the separate contributions of word reading and colour naming depends on the assumption that these two processes are independent. Thus, the purpose of Lindsay and Jacoby's most relevant experiment was to obtain evidence of a dissociation between word reading and colour naming. They reasoned that degrading the colour of Stroop items would reduce the contribution of colour naming, while leaving the contribution of word reading unchanged. There were two colour conditions; bright and dull. Relative to the bright colour condition, in the dull colour condition each of the five colours used in the experiment (black, blue, green, red, and brown) was degraded such that it was harder to distinguish between them. There were an equal number of congruent and incongruent items presented. Control items consisted of a string of symbols (%%%%%) printed in one of the five colours. Participants were required to name the ink colour aloud as quickly as possible within the 800 ms deadline. An experimenter coded the responses as accurate or inaccurate.

The proportion of accurate responses was compared for responses that occurred at speeds between 600 ms and 800 ms. Starting at reaction times (RTs) over 600 ms, the likelihood of making a correct response was plotted at 50 ms intervals (600-649 ms, 650-699 ms, and so on). In the bright colour condition, there was an asymmetrical pattern of interference and facilitation with more interference on incongruent trials than there was facilitation on congruent trials. However, there is the possibility of ceiling effects at the slowest response interval (800 ms). Specifically, at the longest response interval, participants performed equivalently on congruent and control items. Therefore, the rest of the analysis only looked at the fastest response interval (600 ms). Participants were significantly more accurate at naming colours in the bright colour condition than in the dull colour condition, and, again, an asymmetrical pattern of interference and facilitation was apparent. Participants were about 40% less likely to accurately name the colour of incongruent items than of control items, but only about 10% more accurate at naming the colour of congruent items than of control items. In the dull colour condition, however, the amount of interference that occurred on incongruent trials closely matched the amount of facilitation observed on congruent trials (about 25% in both cases). Accurate responses faster than 600 ms were rarely observed on either the incongruent or the control trials and were therefore excluded from the analysis.

Using the formulas described above, Lindsay and Jacoby (1994, Experiment 2) calculated the relative contribution of colour naming and word reading to performance in the two conditions. Colour naming contributed less to performance in the dull colour condition than in the bright colour condition. In addition, degrading the word colour had no affect on the contribution of word reading. These results demonstrate a process-dissociation in that manipulation of colour quality was found to affect colour naming but not word reading. Lindsay and Jacoby therefore interpreted this result as providing support for the independent contribution of colour naming and word reading to performance in the Stroop task.

In a follow-up experiment, Lindsay and Jacoby (1994, Experiment 3) manipulated the proportion of congruent items, taking advantage of the proportion-congruent effect (Logan, 1980), in an attempt to generate additional evidence of the independence of word reading and colour naming. In the 'mostly congruent' condition, participants were presented with 100 congruent items, 20 incongruent items, and 20 control items. In the 'mostly incongruent' condition, participants were presented with 20 congruent items, 100 incongruent items, and 20 control items. Participants were about 14% more likely to make an error on incongruent trials than on control trials in the 'mostly incongruent' condition. In contrast, in the 'mostly congruent' condition, interference was much higher, with the likelihood of making an error 42% greater on incongruent trials than on control trials. Overall, relative to control trials, participants were more accurate on congruent trials, but this advantage was approximately equal for the 'mostly congruent' and 'mostly incongruent' conditions. That is, accuracy on congruent trials was not affected by the probability manipulation.

Calculation of the contributions of word reading and colour naming indicated that the colour naming process was not affected by manipulating the proportion of congruent trials. However, the contribution of word reading was significantly higher in the mostly congruent condition. Lindsay and Jacoby (1994) interpreted their findings as suggesting that participants relied less on the word reading process to generate responses in the mostly incongruent condition than in the mostly congruent condition. In contrast, contribution of the colour naming process was unaffected by the proportion of congruent trials. This process-dissociation provided further support for Lindsay and Jacoby's claim that word reading and colour naming occur simultaneously and contribute independently to performance of the Stroop task

One of the assumptions of Cohen et al.'s (1990) parallel distributed processing model is that the likelihood of encountering incongruent items across the experimental session modulates the relative strength of the colour naming and word reading processes. This modulation occurs on each trial in response to an assessment of the amount of response conflict that originated from word processing. The word reading process will be stronger when the probability of congruent items is high, and weaker when the probability of congruent items is low. From this perspective, the amount of Stroop interference that occurs on any trial is determined by changes in the strength of the word reading and colour naming processes caused by response conflict experienced on all preceding trials.

Jacoby, Lindsay, and Hessels (Experiment 2B, 2003) tested this account in a recent article in which they manipulated the proportion of congruent trials for particular colours. Specifically, the colour words "blue" and "yellow" were presented in a congruent ink colour 80% of the time (Mostly Congruent items), and the words "white" and "green" were presented in an incongruent ink colour 80% of the time (Mostly Incongruent items). Across the experimental session, however, the proportion of incongruent trials was 50%. Control items consisted of a string of coloured symbols (%%%%%). Items were presented one at a time, and participants were required to name the ink colour aloud within 550 ms or the item disappeared and the screen would flash black. The dependent measure was the accuracy with which the ink colour was correctly named within the deadline. Responses over 550 ms were not included in the analysis.

If the account of the proportion-congruent effect offered by Cohen et al. (1990) is correct, then equal interference should be observed for both Mostly Congruent and Mostly Incongruent colour words. Given that specific colour words appeared at random across trials, conscious expectations about what colour word would appear next could not play a role in the contribution of word reading to response selection. That is, following the logic of Cohen et al.'s (1990) parallel distributed processing model, participants should not be sensitive to the item specific congruency manipulation because the overall probability of congruent items was 50%. Participants should only be sensitive to the accumulated probability of response conflict on previous trials. However, contrary to this prediction, Jacoby et al. (2003) observed different amounts of interference for the Mostly Congruent and Mostly Incongruent items. For the Mostly Congruent items, participants were 63% more accurate on congruent trials than on incongruent trials. However, for the Mostly Incongruent items, participants were only 43% more accurate on congruent trials than on incongruent items. These results suggest the parallel distributed processing model as originally formulated is inadequate to explain the proportion congruent effect. Rather than Stroop interference reflecting the weighting of pathways for colour naming and word reading based on the accumulated probability of response conflict on previous trials, it appears that the influence of word reading can be flexibly modulated for specific items.

The Auditory Stroop Effect

As with the visual Stroop effect, the auditory Stroop effect provides a clear demonstration that irrelevant acoustic information may conflict with a primary task. For this type of interference to occur, both relevant and irrelevant dimensions of a stimulus must be processed simultaneously (MacLeod, 1991). Although there has been a long history of studying interference effects in the auditory domain (e.g., Broadbent, 1954; Cherry, 1953, Hall & Blasko, 2005), there have been relatively few demonstrations of an auditory Stroop effect defined as a direct conflict between two acoustic dimensions. The method that has been used to study the auditory Stroop effect is analogous to the method used to study the visual Stroop effect. Thus, investigators of auditory Stroop effects have required participants to respond to some acoustic feature of a spoken word. For example, participants have been required to judge the pitch of a spoken word as either 'high' or 'low', as a function of whether the word itself is either 'high' or 'low' (Hamers & Lambert, 1972). Whereas the meaning of the word corresponds to its pitch on congruent trials, the meaning of the word does not correspond with its pitch on incongruent trials. Incongruous information from the meaning of the spoken word has been shown to interfere with pitch identification (Hamers & Lambert, 1972). This first demonstration of an auditory Stroop effect illustrated that participants were unable to make judgments of the physical dimension of pitch independently of the meaning of the word.

Similarities Between the Visual and Auditory Stroop Effects

Some researchers have argued that auditory and visual Stroop effects operate in a similar manner and that a common underlying mechanism might explain both effects (Green & Barber, 1981). In contrast, Dyer (1973) claimed that the auditory Stroop effect is a different phenomenon than the visual Stroop effect. Specifically, he claimed that the auditory Stroop effect was an artifact of shadowing spoken words. In Dyer's view the difference in performance between the congruent and incongruent trials, such as the results observed by Hamers and Lambert (1972), might simply occur due to participants shadowing the correct response provided by words on congruent trials. Whereas shadowing words on congruent trials would lead to fast and accurate performance, doing so on incongruent trials would lead to an error. In fact, in Hamers and Lambert (1972) experiment, incongruent trials had an error rate of thirty percent, compared to only six percent in the congruent condition. Dyer (1973) interpreted the differential error rates in Hamers and Lambert's (1972) study as supporting the notion that participants were shadowing the spoken words.

Obviously, this tendency to shadow could not have occurred on all trials, since participants did accurately categorize the pitch of spoken words on 70% of incongruent trials. Even so, if participants did shadow the words on some portion of trials, then a response time benefit would be observed on congruent trials. The possibility that the auditory Stroop effect is merely an artifact of shadowing seriously undermines the proposal that a common underlying mechanism is responsible for auditory and visual Stroop effects.

Green and Barber (1981) sought to test Dyer's account by determining whether an auditory Stroop effect could be observed when participants were required to make a judgment about the gender of a speaker's voice. In this study, the words 'man', 'girl', 'mill', and 'game' were recorded when spoken by both a male and a female. In this case, incongruent Stroop stimuli were represented by the words 'man' or 'girl' spoken by a voice from the opposite gender (male versus female). Congruent Stroop stimuli were defined by a match between the meaning of a word and the gender of the voice in which it was spoken (e.g., the word 'girl' spoken in a female voice). 'Pseudo-congruent' trials were defined by a match between the first letter of a control word and the first letter of a response. For example, the word 'mill' spoken in a male voice would be a pseudo-congruent trial because the first letter 'm' matched the first letter of the correct response. 'Pseudo-incongruent' trials were defined by a mismatch between the first letter of a control word and the first letter of a response. For example, the word 'mill' spoken in a female voice would be a pseudo-incongruent trial because the first letter of the word, 'm', matched the first letter of the incorrect response. Responses were made manually by pressing buttons labeled 'man' or 'girl'.

Green and Barber (1981) found that participants were significantly slower (by 59 ms) to respond on incongruent Stroop trials compared with congruent Stroop trials. Thus, for example, responding 'man' when the word 'girl' was spoken by a male voice tended to result in slower responses than making the same response when the word 'man' was spoken by a male voice. This finding, then, replicated the earlier auditory Stroop effect (Hamers & Lambert, 1972). However, faster responding on congruent trials in Green and Barber's (1981) experiment could not rule out Dyer's (1973) proposal that the auditory Stroop effect was an artifact of shadowing on congruent trials. It was still possible that on congruent trials the participants were shadowing the correct response. Therefore, any difference between congruent and incongruent trials could have resulted from speeded responding on congruent trials rather than from interference from conflicting information on incongruent trials.

Unfortunately, Green and Barber's (1981) results are difficult to interpret because they failed to provide any error rates. However, contrary to Dyer's explanation of the auditory Stroop effect, participants were also 26 ms slower to respond on pseudo-incongruent trials than on pseudo-congruent trials. This 'pseudo-Stroop' effect could not have resulted from shadowing. The responses 'mill' and 'game' were not options, so shadowing these responses would not have led to artificially fast correct responses on pseudo-congruent trials. Similar observations of a pseudo-Stroop effect have been observed in the visual Stroop task. For instance, Dalrymple-Alford (1972) found that participants experienced interference when the first letter of a non-colour word distractor matched the first letter of a possible response. Consider the word 'boat' printed in red ink. In this case, the first letter of the word corresponds to the response 'blue', contributing a source of conflict that can slow responses. The fact that both visual and auditory Stroop tasks produce pseudo-Stroop effects provides some evidence that similar mechanisms may underlie them.

Although Green and Barber's (1981) observation of an auditory pseudo-Stroop effect offers some evidence that the auditory Stroop effect shares a common cause with the visual Stroop effect, it does not entirely address Dyer's criticism that previous auditory Stroop effects were an artifact of shadowing. Green and Barber (1981) specifically tested Dyer's account by seeking to prevent any possibility that shadowing of the spoken words could lead to correct responses on congruent trials. This goal was not met by their demonstration of a pseudo-Stroop effect. Green and Barber's (1981) distractor words (mill and game) were analogous to those used by Dalrymple-Alford

(1972) in that they only indirectly mismatched with the response categories (man and girl) because of a similarity in the first letter.

In this experiment, Green and Barber instructed participants to judge the voice in which each word was spoken as belonging to either 'Dave' or 'Joan' rather than to 'man' or 'girl'. With this procedure, slower responses on incongruent trials could not be attributed to shadowing the word on congruent trials because there was no overlap between any of the words and possible responses. The use of the semantically-related category labels reduced the magnitude of the Stroop effect (26 ms compared to 59 ms), but it was still significant. However, the pseudo-Stroop effect was unaffected by the change in labeling, and was not significantly different from the Stroop effect (35 ms compared to 26 ms). Interestingly, Harrison and Boese (1976) observed a similar reduction in the Stroop effect for semantically-related responses using the visual Stroop task. In their study, participants were required to respond 'blood', 'sun', 'grass', 'chocolate', 'sky', or 'coal' for words presented in either red, yellow, green, brown, blue, or black ink respectively. This experiment produced small, but significant, Stroop interference.

Taken together, the experiments by Green and Barber (1981) provide an illustration of the auditory Stroop effect which can not be attributed merely to shadowing on congruent trials. In addition, the possibility that auditory and visual Stroop effects arise from the same mechanism is strengthened by demonstrations of both auditory and visual pseudo-Stroop effects in which slower responses occur when the first letter of a word differs from the first letter of a potential response (Dalrymple-Alford, 1972; Green & Barber, 1981), and by the observation of reduced, but reliable, Stroop effects in both modalities when responses are semantically-related, but not identical, to the conflicting dimension (Green & Barber, 1981; Harrison & Boese, 1976).

Currently, all but one study dealing with the auditory Stroop effect has focused on the conflict between verbal labels and pitch (e.g., high vs. low and man vs. girl). Morgan and Brandt (1989), however, investigated whether an auditory Stroop effect for the acoustic features of pitch, loudness, and duration could be observed. In this experiment, the relationship between these three features of spoken words was varied to produce congruent, incongruent, and neutral items. Morgan and Brandt (1989) replicated previous demonstrations of the auditory Stroop effect using pitch and verbal labels (Green & Barber, 1981, 1983; Hamers & Lambert, 1972, McLain, 1983). In addition, these authors demonstrated the first auditory Stroop effect using loudness and verbal labels. That is, as is the case with the auditory Stroop effect using pitch, participants were unable to make judgments of the physical dimension of loudness (quiet, loud) independently of the meaning of the word (quiet, loud). For the time condition (fast, slow), Morgan and Brandt did not find a significant auditory Stroop effect. However, this nonsignificant finding is somewhat dubious, given that the alpha was set at .0167 and the p value was equal to .025. If more subjects had been recruited, then a significant result at the stated alpha level may have been obtained.

A Nonverbal Auditory Stroop Effect?

All of the previous auditory Stroop studies provided evidence of interference between an incongruent word meaning and a physical dimension. For example, Hamers and Lambert (1972) presented the spoken words 'high' and 'low' in either a high or low pitch voice. Therefore, an incongruent item involved a direct conflict between the word meaning and the pitch of the spoken word. Green and Barber (1981, 1983) combined the words 'man' and 'girl' spoken in either a male or female voice. In that study as well, the meaning of the word (e.g., man) conflicted with the dimension participants were asked to identify on incongruent trials (e.g., female voice). From these studies, one might conclude that conflict between verbal and nonverbal dimensions are special, with participants unable to avoid the influence of one dimension (verbal or nonverbal) when responding to the other.

The current study constitutes an exploration of whether auditory Stroop interference will occur for two nonverbal dimensions. If so, then any account of auditory Stroop effects as caused by verbal processing having privileged status cannot hold. Instead, such a result would suggest that interference may occur whenever the response associated with the irrelevant acoustic feature conflicts with the response associated with a relevant acoustic feature.

Nonverbal examples of Stroop interference have been quite difficult to find in vision and audition. However, when nonverbal examples of Stroop interference do occur, they tend to rely on pseudoverbal symbols rather than words. Observations of spatial Stroop effects (visual) and the Simon Effect (auditory) belong to this category. In an investigation of a spatial Stroop effect, Funes and Lupianez (2003) required their participants to identify the direction in which an arrow pointed (left or right). The arrow could appear on either the left or right half of the computer screen. Participants made their categorization responses more quickly when the direction in which the arrow pointed and the side of the screen on which it was presented were consistent, than when they were inconsistent.

The Simon Effect (Simon & Small, 1969) provides another example of nonverbal auditory interference. In this case, the interference arises from a motor response conflicting with a spatial location. Simon and Small (1969) presented participants with high- or low-pitched tones to either the right or left ear. Categorization of a tone as high- or low-pitched was made with a left or right button press. The results revealed that participants responded more quickly when the sound was presented to the same location as the response button than when the sound's location and the required response conflicted. For example, if high-pitched tones were to be categorized by a left button press, then participants responded more quickly if they were presented in the left ear than if they presented in the right ear.

However, there are no examples of purely nonverbal visual Stroop effects. This is not surprising given that the primitive features in vision include shape, colour, location, and orientation, which all have independent representations so there is no obvious way to observe Stroop effects by combining them. For example, requiring participants to name the colour red when it is presented in the shape of a circle or square will not produce any interference. The irrelevant dimension must be associated with a response that conflicts with the response associated with the relevant dimension. However, the primitive features in audition include location, loudness, pitch, and timbre; the first three of which have similar, high/low representations. Given the primitive features of audition, it is possible to combine two of these dimensions in such a way that they may be incongruent with one another (e.g., a low pitch sound presented at a high location).

Observation of an auditory Stroop effect for two nonverbal dimensions of a sound would suggest that these dimensions are integral. That is, that the frequency and location of a sound are combined preattentively, and that listeners are unable to ignore either dimension. According to Garner's (1974) logic, stimulus dimensions are integral, or inseparable, to the extent that categorization of either dimension is influenced by the irrelevant dimension.

In a series of experiments, Mondor, Zatorre, and Terrio (1998) provided evidence of the integrality of the location and frequency aspects of a sound. For example, in their Experiment 1, participants were presented on each trial with a single tone that they were required to classify according to either its location or its frequency. Participants were only required to differentiate between two frequencies or two locations within any block of trials. In the frequency task, the tones to differentiate could either be similar (e.g., 500 and 535 Hz or 947 and 997 Hz), or dissimilar (e.g., 500 and 947 Hz or 535 and 997 Hz). For example, participants would classify a 500 Hz tone as 'low', and a 535 Hz tone as 'high' in the similar frequency condition. In the dissimilar frequency condition, the participant would classify the 500 Hz tone as 'low' and the 947 Hz tone as 'high'. In the location task, participants classified the location of a tone as either central or peripheral. The two locations the participants were required to differentiate could either be similar (e.g., 30° and 15° left or 15° and 30° right), or dissimilar (e.g., 0° and 45° left or 0° and 45° right). For example, in the location similar condition, a tone presented from 15° left would be classified as 'central' and a tone presented from 30° left would be classified as 'peripheral'. In the location dissimilar condition, a tone presented from 0° would be classified as 'central' and a tone presented from 45° left would be classified as 'peripheral'.

In one half of the trial blocks, the irrelevant dimension also varied across conditions, resulting in four basic trial types for both the frequency and location discrimination tasks: Similar Location/Similar Frequency (e.g., the locations 15° and 35° and the tones at 500 Hz and 535 Hz), Similar Location/Dissimilar Frequency (e.g., the locations 15° and 35° and the tones at 500 Hz and 947 Hz), Dissimilar Location/Similar Frequency (e.g., the locations 0° and 45° and the tones at 500 Hz and 535 Hz), and Dissimilar Location/Dissimilar Frequency (e.g., the locations 0° and 45° and the tones at 500 Hz and 947 Hz). Thus, the difficulty of each task was manipulated by controlling the similarity of sounds on the task-relevant dimension, while there was also either large or small variation of the dimension that was irrelevant to the task.

In each of the remaining blocks of trials, the irrelevant dimension was held constant to provide a Control condition for evaluating the effect on performance of varying the irrelevant dimension in other blocks of the experiment. For instance, in the Similar Frequency-Control condition, when participants classified sounds according to their frequency, tones would be presented at either 500 Hz or 535 Hz and all tones would be presented from a speaker presented at 15° left. Likewise, in the Similar Location-Control condition, when participants classified sounds according to their location, tones would be presented at either 15° or 35° left and all tones would be presented at a frequency of 500 Hz.

Not surprisingly, Mondor, et al. (1998) found that participants were faster at categorizing tones according to their frequency in the Dissimilar-Frequency than in the

Similar-Frequency condition, and that they were faster at categorizing tones according to their location in the Dissimilar-Location than in the Similar-Location condition. This finding revealed that participants had greater difficulty discriminating sounds that were physically similar. Also, participants were faster at categorizing tones in the Control conditions, when there was no variation of the irrelevant dimension, than when the irrelevant dimension varied across trials. Mondor et al. interpreted this observation as evidence that participants were unable to avoid processing the irrelevant dimension. That is, participants appeared unable to exclusively attend to the task-relevant dimension of the sounds. According to Garner's (1974) logic, this finding suggests that location and frequency are integral aspects of sound. On this basis, the authors interpreted their results as evidence that frequency and location information are preattentively combined into an auditory event. Although these authors demonstrated an interference effect, it was not a Stroop effect. The irrelevant dimension in their study was not associated with a response that conflicted with the response associated with the relevant dimension. However, this demonstration does extend previous research by Melara and Marks (1990), which revealed that loudness, pitch and timbre are also integral dimensions of a sound.

In part, this previous research by Mondor, et al. (1998) provides the motivation for the current study. They showed that participants are unable to avoid processing either location or pitch when it is irrelevant. I expect that performance on a frequency classification task will be impaired when sound location is a source of conflicting information. Similarly, performance on a location classification task should be impaired when sound frequency is a source of conflicting information. More specifically, it ought to be possible to demonstrate an auditory Stroop effect based on conflict between the nonverbal dimensions of sound frequency and sound location. Such a demonstration would differ from previous observations of auditory Stroop effects, which have exclusively involved conflict between verbal and nonverbal dimensions of auditory stimuli.

EXPERIMENT 1

The current project is designed to explore the consequences of a conflict between two acoustic features of a sound. Specifically, if an auditory Stroop effect occurs when there is a conflict between the acoustic features of location and frequency, then it would suggest that a possible cause of auditory Stroop effects is the information from the irrelevant dimension conflicting with the information from the relevant dimension. That is, it is possible that one cause of a Stroop effect is a conflict between the response associated with the irrelevant dimension and the response associated with the relevant dimension. However, if an auditory Stroop effect does not occur for a conflict between two physical dimensions, then it would support the dominant view that Stroop effects are based on the inability of participants to avoid processing the meaning of irrelevant words, and this interferes with responding to the relevant dimension. In Experiment 1, participants will judge the location of sounds on one block of trials, and the pitch of sounds on another block of trials. On some trials, the location and frequency of the sound will be congruent (e.g., a high-pitched sound presented from a high speaker), and on other trials, the location and the frequency of the sound will be incongruent (e.g., a high-pitched sound presented from a low speaker). Evidence of a nonverbal auditory Stroop effect will be apparent if participants respond more slowly on incongruent trials than on congruent trials.

Method

Participants

Twenty four participants were recruited from the University of Manitoba's Introduction to Psychology subject pool (10 females and 14 males). Each participant received course credit in exchange for their participation. All participants were required to have normal hearing.

Materials

Computer and Sound System. The experiment was conducted using a Dell Dimension L733R, Pentium 3 computer connected to a 17-inch colour monitor. The E-Prime software system (Psychology Software Tools, Inc., 1999) was used to present stimuli and record responses. The sounds were presented through Altec Lansing ACS340 speakers at approximately 70dB SPL. One speaker was presented at approximately 11° visual angle above the participant's eye level. The second speaker was presented at 11° visual angle below the participant's eye level.

Sounds. Adobe Audition 1.5 (Adobe Systems Incorporated, 2005) software was used to synthesize sounds with a sampling rate of 44,100 Hz. Two sounds, one low-pitched and one high-pitched, were created based on a sine wave. The low-pitched sound (labeled 'low') included a fundamental frequency of 362 Hz plus the first (724 Hz) and second (1086 Hz) harmonics. Relative to the fundamental frequency, the intensity of the harmonics was set to 50% and 25% respectively. The high-pitched sound (labeled 'high') included a fundamental frequency of 732 Hz plus the first (1464 Hz) and second (2196 Hz) harmonics. Relative to the fundamental frequency, the intensity of the harmonics was set to 50% and 25% respectively. The sounds were 120 ms in duration and included 5 ms onset and offset ramps to eliminate onset or offset clicks.

Procedure

At the beginning of the experimental session, participants were presented with each of the high- and low-pitched sounds from both the upper and lower speaker positions in order to familiarize them with the high/low dimension of both acoustic features. On each trial, participants were required to classify a single sound as either high or low according to its frequency or location. Within a given block, participants were only required to classify tones according to one acoustic feature while ignoring the other feature. Specifically, one block of trials involved categorizing the pitch of sounds as either 'high' or 'low'. The other block of trials required categorizing the location of sounds as either originating from the 'high' or 'low' speaker. The order of these blocks was counterbalanced across participants. Before the start of each block, participants received 24 practice trials to familiarize them with the task relevant to that block.

Each practice trial began with a fixation cross ('+') presented at the center of the computer screen for 300 ms, followed by the presentation of a sound for 120 ms. The participant judged the sound as high or low according to either its pitch or its location by pressing either the left or right button on a mouse. The mapping between button responses (left vs. right) and sound categorization (high vs. low) was counterbalanced across participants. Response times were measured as the time between the onset of the sound and the participant's response. After making a response, participants received feedback about the speed and accuracy of their judgment. The subsequent trial began 1000 ms following this feedback. After the practice session for each block, participants performed 96 experimental trials. The experimental trials were identical to the practice trials, except that participants were required to respond within 1000 ms and no feedback as to their speed or accuracy was provided. After finishing the first block of experimental trials, participants received instructions for the next task.

There were four types of trials in this experiment. Sounds were presented from the upper speaker on half of the trials (High Location condition), and from the lower speaker on the other half of trials (Low Location condition). Sounds presented from each of these locations were high-pitched on half of the trials (High Pitch condition) and low-pitched on the remaining half of trials (Low Pitch condition). Thus, on half of the trials, the pitch and the location of the sounds were congruent, with either a high-pitched sound presented from the high speaker or a low-pitched sound presented from the low speaker. On the other half of the trials, the pitch and the location of the sounds were incongruent, with either a high-pitched sound presented from the low speaker or a low-pitched sound presented from the high speaker. Trials corresponding to each combination of sound location and sound pitch (High Location/High Pitch, High Location/Low Pitch, Low Location/High Pitch, and Low Location/Low Pitch) were presented in random order.

Results

Within each cell of the design any observation more than 2.5 standard deviations above or below the mean was eliminated from the analysis and the mean was recalculated. This procedure eliminated less than 2% of the observations from either task. The average mean correct response times (RT) and proportion of incorrect responses for each participant were then submitted to separate 2 x 2 x 2 mixed-design ANOVAs, treating Task Order (pitch task first vs. location task first) as a between-participants factor and Trial Type (congruent vs. incongruent) and Judgment (pitch vs. location) as within-participant factors. Mean RT and error rates for each condition are displayed in Table 1.

Table 1
Mean RTs and Proportion Errors for the Pitch and Location tasks, as a Function of Trial Type in Experiment 1. Standard errors for both measures are shown in parentheses.

		RT	ERRORS
PITCH JUDGMENT			
	Congruent	464 (12.96)	.03 (.01)
	Incongruent	487 (14.50)	.05 (.02)
LOCATION JUDGMENT			
	Congruent	512 (15.52)	.08 (.02)
	Incongruent	550 (19.38)	.15 (.03)

Response Times

Mean RTs on congruent and incongruent trials for the pitch and location judgments are displayed in Figure 1. In the analysis of RTs, there was no significant main effect of Task Order, $F(1, 22) = 1.714, p = .204$. That is, whether participants judged the pitch dimension or the location dimension in the first or second block of the experiment did not influence performance. The analysis did reveal significant main effects of Judgment, $F(1, 22) = 6.549, p < .05$, and Trial Type, $F(1, 22) = 25.836, p < .001$. Participants classified sound pitch about 57 ms faster than they did sound location, and responded more quickly on congruent trials than on incongruent trials. Task Order did not significantly interact with either Trial Type or Judgment ($F < 1$ in both cases). In addition, the interaction between Trial Type and Judgment was not significant, $F(1, 22) = 1.161, p = .293$. The three-way interaction between Task Order, Trial Type, and Judgment also failed to reach statistical significance, $F < 1$.

These findings are consistent with the idea that location and frequency are combined preattentively, and that any variation in one dimension influences classification time based on the other dimension. This finding provides the first demonstration of an auditory Stroop effect owing to a conflict between two non-verbal dimensions. In addition, it provides the first evidence that an auditory Stroop effect may occur due to a conflict between the response associated with the irrelevant dimension and the response associated with the relevant dimension.

Errors

Mean percentage error on congruent and incongruent trials for both the pitch and location judgments are displayed in Figure 2. Analysis of the error data revealed a complementary pattern of performance. The main effect of Task Order was not significant, $F(1, 22) = 3.276, p = .084$. However, the main effects of Judgment, $F(1, 22) = 11.205, p < .01$, and Trial Type, $F(1, 22) = 8.601, p < .01$, were statistically significant. Participants made approximately 7% more errors judging sound location than they did in judging sound pitch, and made fewer errors on congruent trials than on incongruent trials. The interaction between Judgment and Trial Type was not significant, $F(1, 22) = 1.519, p = .231$. Task Order did not significantly interact with either Trial Type, $F(1,22) = 3.367, p = .080$, or Judgment, $F(1,22) = 2.401, p = .136$. The three-way interaction between Task Order, Trial Type, and Judgment was also not statistically significant, $F < 1$.

Figure 1. Mean Reaction Time for Congruent and Incongruent Trials in Experiment 1

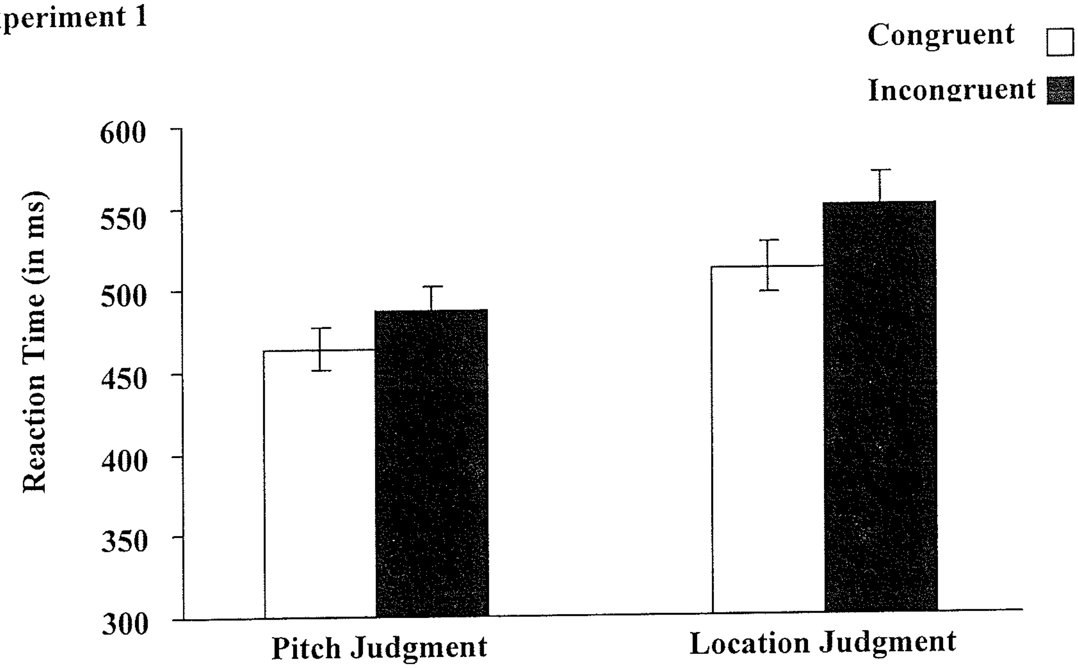
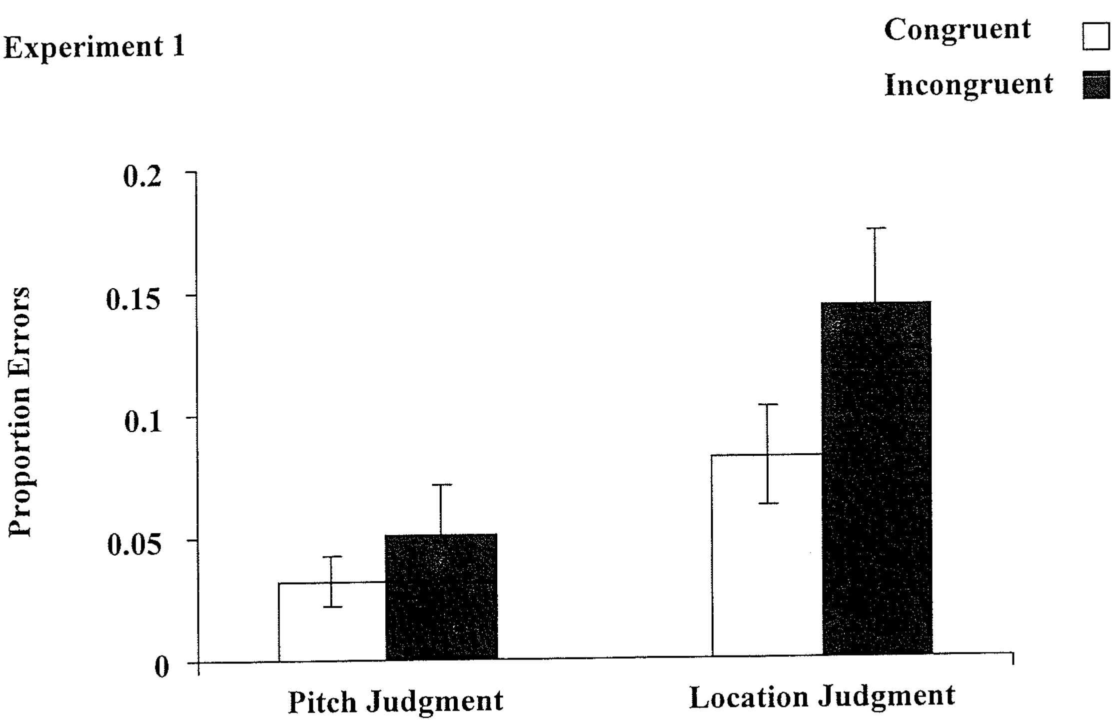


Figure 2. Mean Proportion of Errors for Congruent and Incongruent Trials in Experiment 1



Discussion

These results provide the first demonstration of an auditory Stroop effect owing to a conflict between the responses associated with two nonverbal acoustic features. Existing theoretical accounts of Stroop effects that focus on a conflict between word-reading and colour-naming are clearly incompatible with such a result. For example, the automaticity account of Stroop interference is based on the idea that an automatic process (i.e., word-reading) will interfere with a controlled process (i.e., colour-naming) but not vice versa (MacLeod, 1991). The results of Experiment 1 reveal sizable Stroop interference effects irrespective of whether location or pitch was to be ignored. Given previous evidence suggesting that frequency and location can not be attended separately (Mondor, et al., 1998), an obvious explanation of these interference effects is that both features are processed automatically and that a response is derived for each of them. Some time is required to resolve the incompatibility of these candidate responses on incongruent trials but not, of course, on congruent trials. Thus, the source of auditory Stroop interference may be primarily at the level of the response codes associated with my sounds.

EXPERIMENT 2

Given that Experiment 1 demonstrated that an auditory Stroop effect could be observed for a conflict between two physical dimensions of a sound, Experiment 2 addressed whether the relative influence of the irrelevant dimension could be modulated. Research by Jacoby, Lindsay, and Hessels (2003) on the Item-Specific Congruency

Effect illustrated that with visual Stroop items, participants are able to flexibly modulate the influence of the irrelevant dimension. Specifically, in their study, when one set of colour words was presented mostly in an incongruent colour across the experimental session, the Stroop effect was smaller than for another set of words that were most often presented in a congruent colour. This result suggests that participants can use the likelihood that word identity is predictive of a matching colour identification response as a cue for modulating the contribution of word reading to response selection.

The purpose of Experiment 2 was to investigate whether the auditory Stroop effect observed in Experiment 1 would produce an item-specific congruency effect similar to that observed by Jacoby, et al. (2003). To test this idea, I manipulated the proportion of congruent trials (Congruent Probability variable) for particular pitches and locations. For the frequency judgment task, tones presented to one location were presented in a congruent pitch 80% of the time (Frequency Task, High-Probability Congruent condition), and tones presented to the other location were presented in an incongruent pitch 80% of the time (Frequency Task, Low-Probability Congruent condition). Similarly, for the location judgment task, one tone pitch was presented in a congruent location 80% of the time (Location Task, High-Probability Congruent condition), while tones of the opposite pitch were presented in an incongruent location 80% of the time (Location Task, Low-Probability Congruent condition).

Thus, for both the location and frequency judgment tasks, the proportion of incongruent stimuli was 50% across the experimental session, as in Experiment 1. The only modification from Experiment 1 was that the irrelevant dimension of particular tones was either predictive or non-predictive of a congruent frequency or location response. If there is no significant difference between the High-Probability Congruent and Low-Probability Congruent conditions, then this would suggest that participants can not flexibly modulate the influence of the irrelevant dimension. Such a result would suggest that the nonverbal auditory Stroop effect is caused by a different mechanism than the visual Stroop effect. In contrast, if participants can flexibly modulate the influence of the irrelevant dimension, then the auditory Stroop effect should be smaller in the Low-Probability Congruent condition than in the High-Probability Congruent condition. Such a result would suggest that for nonverbal auditory Stroop items, participants were able to use the likelihood that the irrelevant acoustic feature was predictive of a correct response to modulate its contribution to response selection. Moreover, this result would provide evidence that the relative contribution of each dimension to response selection is flexible, similar to the Item-Specific Congruency Effect observed with visual Stroop items.

Method

Participants

Thirty-two participants were recruited from the University of Manitoba's Introduction to Psychology subject pool (25 females and 7 males). Each participant received course credit in exchange for their participation. All participants were required to have normal hearing.

Materials

The same materials were used as in Experiment 1.

Procedure

Experiment 2 was the same as Experiment 1 except that; 1) the probability that tones of a particular frequency were presented in a congruent location was manipulated for the location task, and 2) the probability that tones presented in a particular location were presented at a congruent frequency was manipulated for the frequency task. For the frequency judgment task, the specific tone location (high vs. low) selected for the High-Probability Congruent and Low-Probability Congruent conditions was counterbalanced across participants. Similarly, for the location judgment task, the specific frequency (high vs. low) assigned to the High-Probability Congruent and Low-Probability Congruent conditions was counterbalanced across participants. Fully counterbalancing combinations of sound location and sound pitch assigned to the High-Probability Congruent and Low-Probability Congruent conditions for each of the two tasks required the generation of four different versions of the experiment (see Table 2 for the sound pitch and sound location assignments to the High-Probability Congruent condition). The number of experimental trials per task was increased to 240 (from 96 in Experiment 1) to accommodate the addition of the Congruent Probability variable to the design. For illustrative purposes, Appendix A displays the number of trials corresponding to each combination of sound location and sound pitch for both the location and frequency judgment tasks for Version 1 of the experiment (see left column of Table 2).

Table 2
Sound pitch and sound location assignments to the High-Probability Congruent condition for the frequency and location judgments.

	VERSION 1	VERSION 2	VERSION 3	VERSION 4
FREQUENCY JUDGMENT	High Location	High Location	Low Location	Low Location
LOCATION JUDGMENT	High Pitch	Low Pitch	High Pitch	Low Pitch

Results

Within each cell of the design any observation more than 2.5 standard deviations above and below the mean was eliminated from the analysis and the mean was recalculated. This procedure eliminated less than 2% of the observations from either task. The average mean correct reaction times (RT) and proportion of incorrect responses for each participant were then submitted to separate 2 x 2 x 2 x 2 mixed-design ANOVAs, treating task order (pitch task first vs. location task first) as a between-participants factor and Probability Congruent (high vs. low), Trial Type (congruent vs. incongruent), and Judgment (pitch, location) as within-participant factors. Mean RT and error rates for each condition are displayed in Table 3.

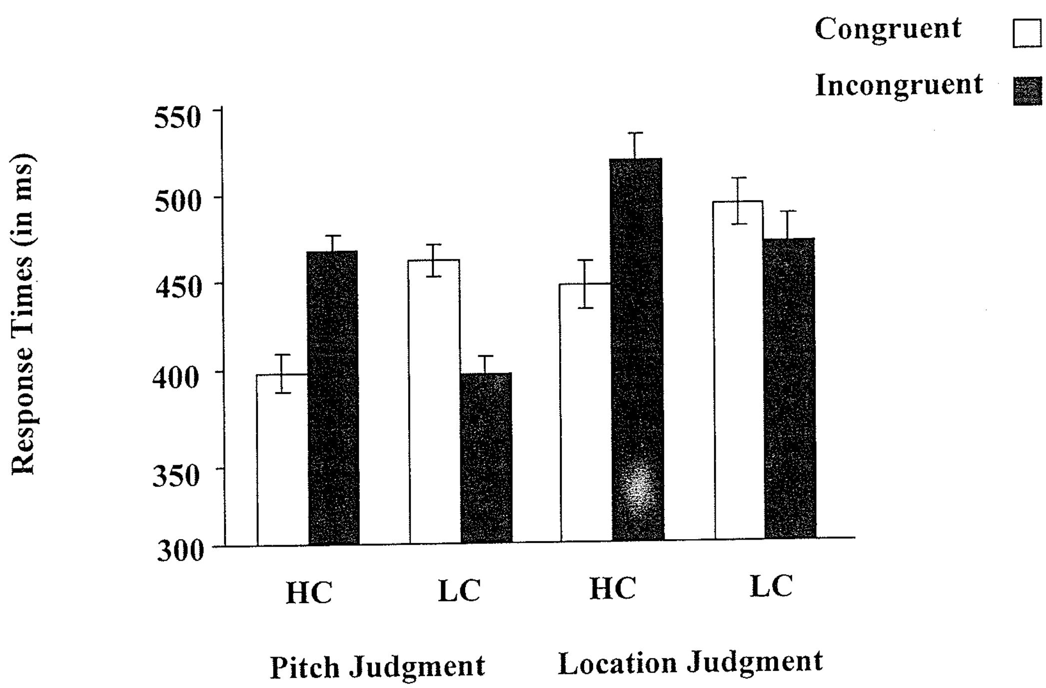
Table 3
Mean Response Times (RTs), Proportion of Errors (ERR) and Standard Errors (SE) for both measures are shown for the Pitch and Location Judgments, as a Function of Probability Congruent and Trial Type in Experiment 2.

	High Probability Congruent		Low Probability Congruent	
Pitch Judgment	RT	Errors	RT	Errors
Congruent	398 (10.67)	.02 (.00)	462 (9.11)	.07 (.01)
Incongruent	467 (9.53)	.09 (.02)	396 (11.21)	.03 (.01)
Location Judgment				
Congruent	447 (13.75)	.05 (.01)	492 (13.46)	.08 (.01)
Incongruent	518 (14.89)	.13 (.02)	471 (15.44)	.11 (.03)

Response Times

Figure 3 displays the mean reaction times for the pitch and location tasks as a function of sound congruency for both the high- and low-probability congruent conditions. Analysis of RTs yielded significant main effects of Judgment, $F(1,30) = 18.82$, $p < .01$, and Trial Type, $F(1, 30) = 12.165$, $p < .01$. Participants classified sound pitch about 51 ms faster than they did sound location, and responded more quickly on congruent trials than on incongruent trials. There was no significant main effect of Task Order, $F(1,30) = 1.148$, $p = .292$, or of Probability Congruent, $F < 1$. That is, whether participants judged the pitch dimension or the location dimension in the first or second block of the experiment did not influence performance on either task as a function of trial type or probability congruent. The main effect of Probability Congruent and the interaction between Probability Congruent and Judgment, were not statistically significant ($F < 1$ in both cases).

Figure 3. Mean Response Time as a function of Probability Congruent and Trial Type in Experiment 2



Note: HC = High-Probability Congruent; LC = Low-Probability Congruent

Although two-way interactions between Trial Type and Task Order, and between Judgment and Task Order, were not statistically significant ($F < 1$, in both cases), interactions between Judgment and Trial Type, $F(1, 30) = 11.529, p < .01$, between Probability Congruent and Trial Type, $F(1,30) = 71.353, p < .001$, and between Probability Congruent and Task Order, $F(1,30) = 6.235, p < .05$, were statistically significant.

Several higher-order interactions were significant as well. The 3-way interaction between Probability Congruent, Trial Type, and Judgment was significant, $F(1,30) = 6.539, p < .05$. None of the other three-way interactions were significant ($p > .26$ in all cases). Most importantly, however, the four way interaction that included all four factors (i.e., Task Order, Probability Congruent, Trial Type, and Judgment) was also significant, $F(1,30) = 5.882, p < .05$. I decomposed this interaction by examining the effect of Probability Congruent, Trial Type and Judgment for each task order separately.

When the Pitch judgment was performed first, main effects of Trial Type, $F(1,15) = 8.299, p < .05$, Judgment, $F(1,15) = 6.844, p < .05$, and Probability Congruent were all significant, $F(1,15) = 4.448, p = .052$. Participants classified sound pitch about 41 ms faster than they did sound location. The interaction between Trial Type and Probability Congruent was also significant, $F(1,15) = 37.498, p < .001$. This interaction arose because participants judged congruent trials approximately 74 ms faster than incongruent trials in the high probability congruent condition, $F(1,15) = 42.838, p < .001$, but judged congruent trials 41 ms more slowly than congruent trials in the low probability congruent condition, $F(1,15) = 14.527, p < .01$. In addition, a significant interaction between Trial Type and Judgment, $F(1,15) = 9.515, p < .01$, arose because whereas there was no significant difference in responding to congruent or incongruent trials when judging pitch, $F < 1$, participants responded to congruent trials approximately 31 ms faster than incongruent trials when they judged location, $F(1,15) = 21.431, p < .001$. Neither the two-way interaction between Probability Congruent and Judgment was significant, $F < 1$, nor the three-way interaction between Trial Type, Probability Congruent, and Judgment, $F(1,15) = 1.532, p = .235$, were significant.

Performance when location was judged first was slightly different. Whereas the main effect of Judgment was significant, $F(1,15) = 12.065, p < .01$, and the main effect of Trial Type, $F(1,15) = 3.985, p = .064$, approached significance, the main effect of

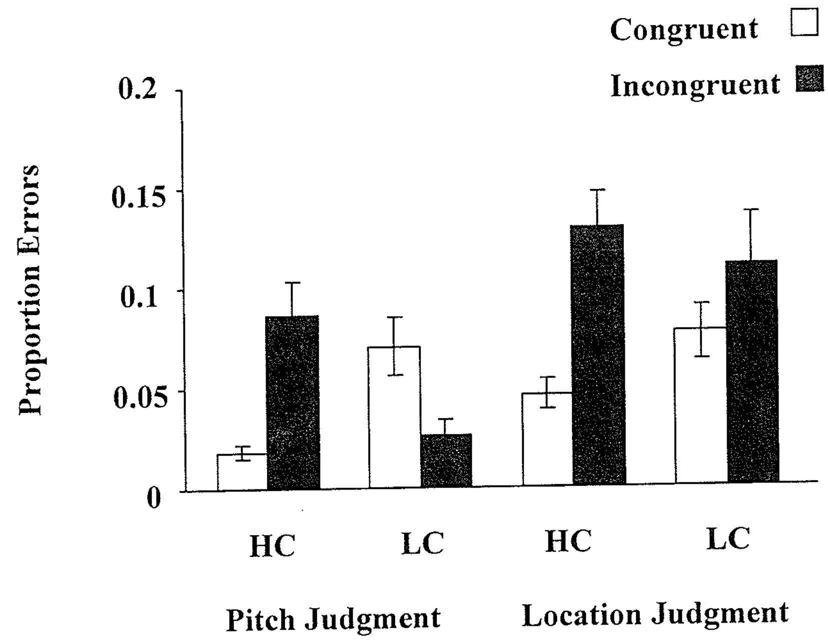
Probability Congruent, $F(1,15) = 1.889, p = .189$, was not significant. The two-way interactions between Probability Congruent and Judgment, and between Trial Type and Judgment were not statistically significant ($p > .11$ in both cases). However, both the two-way interaction between Trial Type and Probability Congruent, $F(1,15) = 33.944, p < .001$, and the three-way interaction between Trial Type, Probability Congruent, and Judgment, $F(1,15) = 15.938, p < .001$, were statistically significant. I decomposed this interaction by examining the effects of Trial Type and Judgment for the Probability Congruent conditions separately. In the High Probability Congruent condition, the main effects of Trial Type and Judgment were statistically significant ($p < .01$ in both cases), but the interaction between these factors was not, $F(1,15) = 3.239, p = .092$. Participants judged a sound's pitch approximately 62 ms faster than a sound's location, and were approximately 65 ms faster on congruent trials than on incongruent trials. In the Low Probability Congruent Condition, the main effects of Trial Type, $F(1,15) = 20.598, p < .001$, and Judgment, $F(1,15) = 10.348, p < .01$, were significant. Participants judged a sound's pitch approximately 62 ms faster than a sound's location, however, unlike the High Probability Congruent condition, participants were approximately 46 ms faster on incongruent trials than on congruent trials In addition, the two way interaction between these factors (Trial Type and Judgment) was significant, $F(1,15) = 19.252, p < .001$. I decomposed this interaction by examining the effect of Trial Type in each Judgment separately. In the Pitch Judgment, there was a significant main effect of Trial Type, $F(1,15) = 37.636, p < .001$. Participants were approximately 25 ms faster on incongruent trials than on congruent trials. In the Location Judgment, there was no significant main effect of Trial Type, $F(1,15) = 2.348, p = .146$.

Errors

Figure 4 displays the proportion of errors for the pitch and location tasks as a function of sound congruency for both the high- and low-probability congruent conditions.

In the analysis of the error data, there was a main effect of Judgment, $F(1,30) = 12.308, p < .05$, and Trial Type, $F(1,30) = 11.755, p < .01$. There was no significant main effect based on Task Order or Probability Congruent ($F < 1$ in both cases). Thus, whether participants judged the pitch dimension or the location dimension in the first or second block of the experiment did not influence performance on either task. In addition, whether participants judged sounds in the High- or Low-Probability Congruent conditions did not influence performance, $F < 1$.

Figure 4. Mean Proportion of Errors as a Function of Probability Congruent and Trial Type in Experiment 2



Note: HC = High-Probability Congruent; LC = Low-Probability Congruent

Further analysis revealed that there were significant 2-way interactions between Judgment and Trial Type, $F(1,30) = 6.103, p < .05$, and between Probability Congruent and Trial Type, $F(1,30) = 18.752, p < .001$. None of the other 2-way interactions were significant ($p > .14$ in all cases). The 3-way interaction between Judgment, Trial Type and Probability Congruent approached significance, $F(1,30) = 3.520, p =.07$. I decomposed this interaction by examining the effects of Trial Type and Judgment for the Probability Congruent conditions separately. In the High Probability Congruent condition, the main effects of Trial Type, $F(1,31) = 27.504, p < .001$, and Judgment,

$F(1,31) = 11.192$, $p < .01$, were significant. Participants judged a sound's pitch approximately 3% more accurately than a sound's location, and were approximately 7% more accurate on congruent trials than on incongruent trials. In the High Probability Congruent condition, the interaction between Trial Type and Judgment was not significant, $F < 1$. In the Low Probability Congruent Condition, the main effect of Trial Type was not significant, $F < 1$, but the main effect of Judgment was significant, $F(1,31) = 9.050$, $p < .01$. In addition, the two way interaction between these factors (Trial Type and Judgment) was significant, $F(1,31) = 6.046$, $p < .05$. I decomposed this interaction by examining the effect of Trial Type in each Judgment separately. In the Pitch Judgment, there was a significant main effect of Trial Type, $F(1,31) = 16.696$, $p < .001$. Participants were approximately 5% more accurate judging the pitch dimension on incongruent trials than on congruent trials. In the Location Judgment, there was no significant main effect of Trial Type, $F(1,31) = 1.513$, $p = .228$.

None of the other 3-way interactions were statistically significant ($F < 1$ in all cases).

Discussion

The results of Experiment 2 provide a clear and unequivocal demonstration that the extent to which incongruent response information interferes with performance of a primary task may vary markedly. Under conditions in which a mismatch is quite likely (the low probability congruent condition), responses are actually executed more quickly on incongruent trials than on congruent trials. In contrast, under conditions in which a mismatch is quite unlikely (the high probability condition), responses are executed more quickly on congruent trials than on incongruent trials. Thus, not only the magnitude, but also the direction of the Stroop interference effect, is influenced by the probability that for a specific item the responses associated with the judged and the to-be-ignored features are compatible. Intriguingly, this pronounced effect differs markedly from previous demonstrations of item-specific congruency effects in vision (Jacoby, et al., 2003) in which only the magnitude of Stroop interference was influenced.

General Discussion

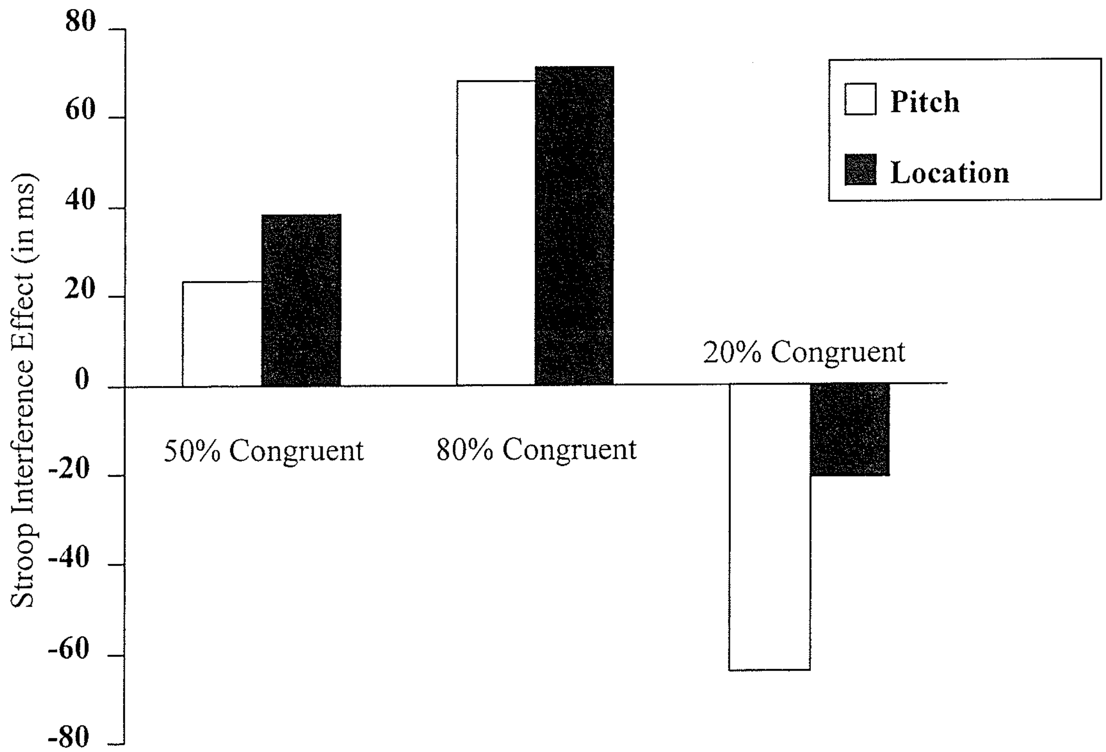
Current explanations of Stroop interference are designed to explain conflict in responding to some physical dimension of a word based on inconsistent information provided by the word's meaning (Cohen, Dunbar, & McClelland, 1990; Green & Barber, 1981, 1983; Hamers & Lambert, 1972; Jacoby et al., 2003; Lindsay & Jacoby, 1994; Stroop, 1935). For example, according to Lindsay and Jacoby (1994) both interference and facilitation occur in the Stroop paradigm due to the independent and simultaneous processing of word reading and colour naming. Thus, Lindsay and Jacoby (1994) argued that "the Stroop effect is the paradigmatic example of situations in which two types of cognitive processes, one intended and the other automatic, simultaneously contribute to performance" (p. 233).

Experiment 1 of the present study provides evidence of an auditory Stroop effect based on a conflict between responses associated with two nonverbal, primitive acoustic features. This finding represents a novel variant of the Stroop effect in the auditory modality. Moreover, there appear to have been no analogous reports of a visual Stroop effect based on a conflict between two primitive visual features (see, for example, MacLeod, 1991 for a review).

As shown in Figure 5, in Experiment 2 I found that the degree to which an irrelevant feature interferes with performance depends on whether that feature is predictive of a congruent or incongruent response to the task-relevant dimension. When Jacoby et al. observed a similar item-specific congruency effect with the visual Stroop task, their interpretation was that "...early processing of individual words triggers inhibitory processes that curtail full reading of the word or block access of any word-reading processes to the response system." (Jacoby, et al., 2003, p. 643). In this view, making the irrelevant dimension unlikely to correspond to the required response causes an attenuation of its contribution to the generation of color identification responses. Nevertheless, this explanation is not consistent with my observation that participants were faster at responding to incongruent than congruent sounds in the low-probability congruent condition.

An obvious and simple account of the results of Experiment 1 is that on-line algorithmic processing of each sound yields two responses, one associated with location and one associated with pitch. A response is executed more quickly when these responses are consistent than when they are inconsistent. Thus, conflict between the responses associated with location and pitch delay responses on incongruent trials relative to congruent trials. The results of Experiment 2, however, may not be explained in this way because performance did not depend solely on the congruency of the responses associated with the pitch and location of sounds. Rather, performance was best for trial types that were experienced most often, irrespective of the compatibility of the responses associated with the location and frequency of a sound.

Figure 5. Magnitude of the Stroop interference effect as a function of the Probability Congruent trials in Experiments 1 and 2



Note: '50% congruent' describes performance in Experiment 1. '80% congruent' describes performance in the high-probability congruent condition of Experiment 2. '20% congruent' describes performance in the low-probability congruent condition of Experiment 2. Positive values represent interference caused by the irrelevant feature and negative values indicate facilitation caused by the irrelevant feature.

One possible account of these results is based on an interaction between online, algorithmic processing of each sound and retrieval of similar 'instances' stored in memory (Jacoby, 1978; Logan, 1988, 2002). These accounts assume that the likelihood and speed of retrieving a memory episode to facilitate generation of a response to a new stimulus depends on the frequency of having done so in the past (Hintzman, 1976; Jacoby & Brooks, 1984; Logan, 1988). According to Logan's instance theory, "… novices begin with a general algorithm that is sufficient to perform the task. As they gain experience, they learn specific solutions to specific problems, which they retrieve when they encounter the same problems again. Then, they can respond with the solution retrieved from memory or the one computed by the algorithm" (Logan, 1988, p. 493). Logan's theory is founded on assumptions that; any attended event forms an episodic trace in memory, present circumstances automatically cause retrieval of representations for similar events stored in memory, and representations for each prior instance are stored independently. Although current events will automatically retrieve similar instances, these instances will not necessarily guide response generation. According to Logan, "the algorithm, if used in parallel with retrieval, will screen out any slow or difficult retrievals by finishing first and providing a solution to the task" (Logan, 1988, p. 494).

It appears that Logan's instance theory may provide a simple explanation of the results of my second experiment. In Experiment 2, the relative frequency of prior responses to specific sounds would determine the likelihood of generating a fast response based on the retrieval of one or more instances of having responded to that sound previously. If fewer memory episodes are available of having responded to specific sounds on previous trials, then there would be a greater likelihood that participants would arrive at a response based on the slower algorithmic process. Participants responded to four times as many congruent as incongruent sounds in the high-probability congruent condition and four times as many incongruent as congruent sounds in the low-probability congruent condition. From the perspective of Logan's instance theory, then, the predictable outcome was that participants were fastest at responding to congruent sounds in the high-probability congruent condition, yet were faster at responding to incongruent sounds in the low-probability congruent condition.

The relative contribution of memory retrieval to the generation of pitch and location responses appears to provide a straightforward explanation for the results of Experiment 2. However, a difference between congruent and incongruent trials in the availability of prior instances cannot account for the auditory Stroop effect observed in Experiment 1 because participants responded equally often to all possible combinations of sound pitch and sound location. Thus, the nonverbal Stroop interference apparent in Experiment 1 may well reflect a difference between congruent and incongruent trials in the efficiency of the on-line algorithmic processing. Assuming that location and pitch are processed simultaneously (Mondor, et al., 1998), I suggest that participants first applied an algorithm to determine the response associated with each dimension, and then determined which response belonged to the relevant dimension. Because the response associated with both dimensions on congruent trials was the same, participants could simply output that response without engaging in further diagnostic processing. In contrast, algorithmic processing would lead to the generation of conflicting responses on incongruent trials, and this inconsistency would have to be resolved before the correct response could be executed. Thus, responding accurately on these trials required participants to perform the additional step of evaluating which response originated from the task-relevant dimension (see MacLeod, 1998; MacLeod & MacDonald, 2000, for a similar account of the Stroop effect in the visual domain).

Regardless of whether or not a model such as the elementary one outlined above is eventually confirmed, the present results have established the existence of a nonverbal auditory Stroop interference, and shown that its magnitude and direction depend on the probability that responses associated with judged and to-be-ignored features are compatible. These results suggest important differences between the generality and nature of Stroop interference effects in responding to auditory and visual events.

References

Adobe Systems Incorporated (2005). *Adobe Audition 1.5*. San Jose, CA.

Barber, P. (1988). *Applied Cognitive Psychology*. London: Methuen.

Besner, D., & Stolz, J. A. (1999). Unconsciously controlled processing: The Stroop effect reconsidered. *Psychonomic Bulletin & Review*, 6(3), 449-455.

Bookbinder, J., & Osman, E. (1979). Attentional strategies in dichotic listening. *Memory and Cognition*, 7(6), 511-520.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 47, 191-196.

Broadbent, D. E. (1958). *Perception and Communication*. Oxford University Press, New York.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11, 63-65.

Cherry, C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975-979.

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

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125(2), 159-180.

Danziger, S., Estevez, A. F., & Mari-Beffa, P. (2002). Stroop interference effects in partially colored Stroop words. *Psychonomic Bulletin & Review*, 9(3), 536-541.

Dalrymple-Alfrod, E. C. (1972). Sound similarity and color-word interference in the Stroop test. *Psychonomic Science*, 28, 209-210.

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

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1(2), 106-120.

Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review*, 2, 145-173.

Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica,* 30, 96-103.

Funes, M. J., & Lupianez, J. (2003). Posner's theory of attention: A task to measure the attentional functions of orienting, alerting and cognitive control and the interactions between them. *Psicothema*, 15(2), 260-266.

Garner, W. R. (1974). *The Processing of Information and Structure*. Hillsdale, NJ: Erlbaum.

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

Green, E. J., & Barber, P. J. (1983). Interference effects in an auditory Stroop task: Congruence and correspondence. *Acta Psychologica*, 53, 183-194.

Hall, M. D., & Blasko, D. G. (2005). Attentional interference in judgments of musical timbre: Individual differences in working memory. *The Journal of General Psychology*, 132, 94-112.

Hamers, J. F., & Lambert, W. E. (1972). Bilingual interdependencies in auditory perception. *Journal of Verbal Learning and Verbal Behaviour*, 11, 303-310.

Harrison, N. S., & Boese, E. (1976). The locus of semantic interference in the Stroop color-naming task. *Perception & Psychophysics*, 20, 408-412.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-91). New York: Academic Press.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, 17, 649-667.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept formation. *Psychology of Learning & Motivation*, 18, 1-47.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, 10(3), 638-644.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, Englewood Cliffs, NJ.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.

Knipling, R. R., & Wang, J. S. (1994). *Crashes and Fatalities Related to Driver Drowsiness/Fatigue.* Washington, D. C.: U. S. Department of Transportation, National Highway Traffic Safety Administration, Office of Crash Avoidance Research, Research & Development.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.

Lieberman, H. R., Tharion, W. J., Shukitt-Hale, B., Speckman, K. L., & Tulley, R. (2002). Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U. S. Navy SEAL training. *Psychopharmacology*, 164(3), 250-261.

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

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109, 376-400.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, 7, 166-174.

Lowe, D., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology*, 36, 684-700.

Long, J. (1977). Division of attention between simultaneous and successive nonverbal signals varying in discriminability. *Quarterly Journal of Experimental Psychology*, 29(1), 7-30.

Lowe, D. G. (1979). Strategies, context, and the mechanism of response inhibition. *Memory & Cognition*, 7, 382-289.

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

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, 26, 201-211.

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*, 10, 304-315.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4, 383-391.

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin*, 118, 35-54.

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

Mitchell, M. C. (1985). Alcohol-induce impairment of central nervous system function: Behavioral skills involved in driving. *Journal of Studies on Alcohol*, 10, 109-116.

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance,* 24(1), 66-79.

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

Morgan, A. L. R., & Brandt, J. F. (1989). An auditory Stroop effect for pitch, loudness, and time. *Brain and Language*, 36, 592-603.

Moskowitz, H., & McGlothlin, W. (1974). Effects of marijuana on auditory signal detection. *Psychopharmacologia*, 40(2), 137-145.

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The Nature of Remembering: Essays in Honour of Robert G. Crowder* (pp. 69-93). Washington, DC: American Psychological Association.

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

Neill, W. T., & Mathis, K. M. (1998). Transfer-inappropriate processing: Negative priming and related phenomena. In Medin, D. L. (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 38. pp. 1-44). San Diego, CA, US: Academic Press.

Pape, A.M., Wiegmann, D.A., & Shappell, S. (2001). Air traffic control (ATC) related accidents and incidents: A human factor analysis. Presented at the *11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Pilcher, J. J., & Walters, A. S. (1997). How sleep deprivation affects psychological variables related to college students' cognitive performance. *Journal of American College Health*, 46(3), 121-126.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.

Poulton, E. G. (1953). Two-channel listening. *Journal of Experimental Psychology*, 46, 91-96.

Pounds, J., Scarborough, A. & Shappell, S. (2000). A human factors analysis of air traffic control operational errors. *Aviation, Space and Environmental Medicine*, 71, 329.

Psychology Software Tools, Inc. (1999). *E-Prime Software System*. Pittsburg, PA.

Rantanen, E. M., & Levinthal, B. R. (2005). Effect of air traffic controller task load and temporal awareness on task prioritization. Presented at the 13th International Symposium on Aviation Psychology, Dayton, OH, 2005.

Rinder, R. J. (1974). Mechanism of input selection in selective perceptual processing of the accepted message in a dichotic auditory presentation. *Journal of Experimental Psychology*, 103(4), 805-807.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic processing: I: Detection, search, and attention. *Psychological Review*, 84, 1-66.

Shappell, S., & Wiegmann, D. (1996). U.S. Naval Aviation mishaps 1977-92: Differences between single- and dual-piloted aircraft. *Aviation, Space and Environmental Medicine*, 66, 65-69.

Shappell, S., & Wiegmann, D. (2000). The human factors analysis and classification system (HFACS) (Report Number DOT/FAA/AM-00/7). Federal Aviation Administration, Washington DC.

Shor, R. E. (1975). An auditory analog of the Stroop test. *Journal of General Psychology*, 93, 281-288.

Simon, J. R., & Small, A. M., Jr. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 53, 433-435.

Spieth, W., Curtis, J. F., & Webster, J. C. (1954). Responding to one of two simultaneous messages. *Journal of Acoustical Society of America*, 26, 391-396.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual task studies of simulated driving and conversing on a cellular phone. *Psychological Science*, 12, 462-466.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, 9, 23-23.

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

Taylor, D. J., & McFatter, R. M. (2003). Cognitive performance after sleep deprivation: Does personality make a difference? *Personality and Individual Differences*, 34(7), 1179-1193.

Thomasius, R., Petersen, K., Buchert, R., Andresen, B., Zapletalova, P., Wartberg, L., Nebeling, B., & Schmoldt, A. (2003). Mood, cognition and serotonin transporter availability in current and former ecstasy (MDMA) users. *Pscyopharmacology*, 167(1), 85-96.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, 37A, 571-590.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology,* 54A, 321-343.

Tipper, S. P., & Cranston, M. (1985) Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology*, 37A, 591-611.

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

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, 20, 727-735.

Webster, J. C., & Thompson, P. O. (1954). Responding to both of two overlapping messages. *Journal of the Acoustical Society of America*, 26, 396-402.

West, R. (1999). Age differences in lapses of intention in the Stroop task. *Journals of Gerontology: Series B. Psychological Sciences & Social Sciences,* 54, P34-P43.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's own name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 255-260.

Wood, T. J., & Milliken, B. (1998). Negative priming without ignoring. *Psychonomic Bulletin & Review,* 5(3), 470-475.

Appendix A: Number of Experimental Trials Corresponding to Each Combination of Sound Pitch and Sound Location for Both the Location and Frequency Judgments for Version 1 of Experiment 2

Location Task

	High Location	**Low Location**	**Total**
High Frequency *(High-Probability Congruent)*	96	24	120
Low Frequency *(Low-Probability Congruent)*	96	24	120
Total	192	48	240

Pitch Task

	High Frequency	**Low Frequency**	**Total**
High Location *(High-Probability Congruent)*	96	24	120
Low Location *(Low-Probability Congruent)*	96	24	120
Total	192	48	240