

Chapter 2

The Behavioral-Related Conditioned Sound in Cross Modal Stroop Task

Hau Lam Agnes Chan and De-Hui Ruth Zhou

Abstract This study examined the effect of conditioned sound as stimuli in an audiovisual Stroop task. It is known that the conditioned sound is induced under the pairing of a sound stimulus with conditioned behaviors in our daily life. For example, the sound of pedestrian traffic light is conditioned with the meaning of crossing the road or not crossing the road. In this study, a cross modal audiovisual Stroop task was used to test the effect of conditioned sound, since it is always presented with other visual cues in our daily life. It was hypothesized that conditioned sound would produce interference effect in the Stroop task. A 2 (audio stimuli: conditioned sound/voiced representation) $\times 2$ (visual stimuli: pictorial representation/text representation) $\times 2$ (congruence: audio stimuli and visual stimuli are congruent in meaning/audio stimuli and visual stimuli are incongruent in meaning) $\times 2$ (tasks: answer corresponding to visual stimuli/answer corresponding to audio stimuli) factorial design was constructed. Sixty Hong Kong college students with permanent Hong Kong resident status participated in this study. ANOVA tests had been employed to explore and analyze the data, aiming to show that both audio stimuli (conditioned sound and voiced representation) might induce interference effect; while voiced representation created greater interference than conditioned sound. The results were in congruence with the previous literature on the induction of behavioral interference under the Stroop task. The study showed that conditioned sound resulted from our daily conditioned behavior affected human perception in the multimodal Stroop effect. The study also supported the attenuation model in selective attention. As conditioned sound may give slower and more inaccurate responses than spoken word, it is suggested that the use of conditioned sound as indicators in different instruments and tools used in our daily life can be

H.L.A. Chan (✉) · D.-H.R. Zhou

Department of Counselling and Psychology, Hong Kong Shue Yan University,
North Point, Hong Kong, China
e-mail: agnes_chl@yahoo.com.hk

D.-H.R. Zhou

e-mail: dhzhou@hksyu.edu

replaced by spoken word to enhance their effectiveness by facilitating quicker and more accurate responses.

Keywords Audiovisual Stroop test • Cross modal • Automaticity • Conditioned sound

2.1 Introduction

There are many signals consisting of sound signal and other signals for human perception in our daily life, for example, the sound signal of the pedestrian traffic light to help the blind people, the sound signal produced by the computer when an USB is inserted or removed to indicate the completion of the process, etc. After familiarizing to these sound signals, which had no meaning attached to them originally, people used to perceive these sounds with their particular meanings. By pairing these sounds with particular behavioral meaning, this kind of linkage formation between a sound and a behavioral-related meaning is called behavioral conditioning (Settlage and Harlow 1936). This study used the term conditioned sound to relate these sounds with a specific conditioned behavioral meaning. As it is questionable that the process of meaning retrieval by conditioned sound is the same as spoken word; this study is going to compare these two sounds in a Stroop test.

2.1.1 *The Stroop Effect and Its Variations*

There are a number of studies on the Stroop effect and its variations in the past 80 years (e.g., Acosta and Simon 1976; Dalrymple-Alford 1968; Jensen and Rohwer 1966; Simon et al. 1975; Stroop 1935). The standard Stroop test is an effect of color word interference (Stroop 1935), which involved a set of color words printed with different color-ink. Under the congruent condition, the color ink of the word had the same meaning as the word did (e.g., the word “blue” was printed with blue-ink; the word “crimson” was printed with red-ink). Under the incongruent condition, the color ink of the word was different from the meaning of the word (e.g., the word “blue” was printed with yellow-ink; the word “crimson” was printed with purple-ink). Reaction time was measured when the participants were asked to read the words or name the color-ink of the words separately. The results showed that the response time under the congruent condition was shorter than that under the incongruent condition; and under the incongruent condition, the response time for reading the words was shorter than that of naming the color-ink.

Interference is the main process leading to the result of the Stroop effect. Interference effect occurs when there is conflict between two or more perceptual

or cognitive processes, where additional time is needed to solve the conflict, causing negative performance impact (Lidwell et al. 2010). Several studies suggested that interference effect was not involved in the encoding stage (Acosta and Simon 1976; Simon 1982; Simon and Berbaum 1990; Simon and Pouraghabagher 1978), but in the response selection stage (Acosta and Simon 1976; Simon 1982; Simon et al. 1975; Van der Molen and Keuss 1981). Simon and Berbaum (1988) suggested that interference occurred between the process of perceptual encoding and response activation, during the process of decoding and information retrieval from short-term memory, which Seymour (1977) mentioned these stages as the conceptual encoding process. In the standard Stroop test, the incongruence of the target (the color-ink of the printed word) and the distractor (the meaning of the printed word) caused conflict between their information retrieval processes which led to interference. Therefore, having shorter response time under the congruent condition than that under the incongruent condition was due to the interference effect.

In addition, automaticity is also observed in the Stroop effect. Automaticity in the Stroop tests refers to the forming of automatic process for reading (i.e., the printed word), brought by long practice and does not need controlled attention (consciously paid attention) to occur (Stirling 1979). Although no controlled attention was needed in the automatic reading process, it still used up part of the attentional resources, thus reduced the resources available to process and name the target (i.e., the color-ink of the printed word; Stirling 1979).

As both the color word and the color-ink in the standard Stroop test required visual sensation to perceive, this kind of Stroop tests are called the visual Stroop tests. Similar visual Stroop effect studies had also shown similar results (e.g., Dalrymple-Alford 1968; Jensen and Rohwer 1966).

Auditory Stroop tests are a kind of variation from the standard Stroop test. The difference between visual Stroop tests and auditory Stroop tests is that the stimuli of visual Stroop tests were visual stimuli (words, pictures, etc.), whereas that of auditory Stroop tests were audio stimuli (sounds, spoken words, etc.). Auditory Stroop tests also got similar results with standard Stroop test (e.g., Green and Barber 1981; Gregg and Purdy 2007; Jerger et al. 1988).

Another kind of variation form of the standard Stroop test is the cross modal audiovisual Stroop tests. It involved both auditory and visual stimuli in the test. Limited studies had investigated this cross modal audiovisual Stroop effects, and Peng and Prime (2014) was the only study that investigated it in both directions (i.e., having both visual distractor and auditory distractor) by a color-naming task. The test involved a set of color words (visual stimulus), a set of color patches (visual stimulus), and a set of sound tract with spoken color words (audio stimulus). Reaction time was measured when participants were asked to read the words (with auditory distractor), named the color patches (with auditory distractor) and read the color word they heard (with visual distractor) separately. The response time under the congruent condition (i.e., the spoken word was congruent with the presented color word or color patches) was shorter than that under the incongruent condition (i.e., the spoken word was different with the presented color word or color patches) due to the interference effect. In addition, under the incongruent

conditions, the response time for the participants to read out the color word they heard (with visual distractor) was slower than that of reading out the words or naming the color patches they saw (with auditory distractor). This showed that visual distraction might exert greater cross modal Stroop interference than audio distraction. To sum up, their study showed that Stroop effects between visual and audio stimuli were multidirectional and visual distractors exerts stronger cross modal Stroop interference (Peng and Prime 2014).

2.1.2 Audiovisual Interference

Apart from those audiovisual studies for the Stroop effects, the audiovisual word interference task was also quite common in the multimodal studies of audiovisual interaction (Calvert et al. 2000; Laurienti et al. 2004; McGurk and MacDonald 1976; Molholm et al. 2002; Soto-Faraco et al. 2004; Van Atteveldt et al. 2004). For instance, McGurk and MacDonald (1976) had reported that visual distraction created by observing the speaker's lip movement might affect the perception of the auditory phonemes. This showed that the audiovisual interaction do really interact with each other, and may then affect the perception of information. On the other hand, the behavioral studies showed that the congruency effect of audiovisual cues might affect behaviors, that is, under the congruent condition (i.e., with the cue of listening to "red" while watching a red circle), the behavioral response would be made faster than normal (i.e., without any cues), whereas under the incongruent condition, the behavior would be slower than normal (Laurienti et al. 2004). This further supported the congruency effect in the audiovisual multimodality (Schriefers and Meyer 1990).

2.1.3 Sound Stimuli: The Conditioned Sound Versus Spoken Word

Auditory Stroop tests which involved audio stimuli (e.g., Cohen and Martin 1975; Green and Barber 1981; Gregg and Purdy 2007; Hamers and Lambert 1972; Jerger et al. 1988), had never used conditioned sound as a stimulus before. A conditioned sound refers to a tone which is conditioned with a meaning. For examples, different traffic light sounds are conditioned with the meaning of "crossing" or "not crossing"; ringing of doorbell is conditioned with the meaning of someone's arrival. Conditioned sound was usually used in the studies of behavioral conditioning (e.g., Imani et al. 2012; Settlege and Harlow 1936). Studies of audio Stroop effects, however, only focused on gender, pitch and color, and did not investigate the effect of conditioned sound before.

The conditioned sound, when compared with a spoken word as a stimulus that is well-studied, has a great difference in the format of meaning.

Listening to spoken words, in the nature of speech, could be a type of variation of reading automaticity as they are comparable in nature. They are both the only way of perceiving semantic information but different towards the type of sensation as reading perceives visual semantic information, whereas listening perceives auditory information. This study assumed that automaticity effect could occur in both reading and listening.

The conditioned sound is not the same as a spoken word, which is originally a meaningful sound, but a meaningless sound paired with a specific meaning over a long time of exposure. Hence, it is hypothesized that the perception process could be different between perceiving spoken word, which is assumed to be automatically listened and its meaning is directly perceived without requiring controlled attention, and perceiving conditioned sound, which involves controlled attention.

2.1.4 Visual Stimuli: Picture Versus Written Words

Several studies had been carried out on the effect of pictures, instead of colors, combined with an embedded written word, as the visual stimuli in visual Stroop tests (e.g., Cattell 1886; Dyer 1973; Fraisse 1969; Golinkoff and Rosinski 1976; Rosinski 1977; Rosinski et al. 1975). The result of the incongruent condition (i.e., meaning of the picture and the embedded word is different) showed that the distraction of written word greatly interfered with the meaning of the picture, but there was only little distraction brought by the picture to the written word when perceived (e.g., Golinkoff and Rosinski 1976; Rosinski 1977; Rosinski et al. 1975).

As a stimulus, picture is similar to the conditioned sound. It does not involve the automaticity effect as written word does, and is paired with a meaning. This study used picture as one of the independent visual stimulus in order to investigate the effect of automaticity.

2.1.5 Key Press Response in Stroop Tests

The standard Stroop test and many of its variations had used oral response for measurement (e.g., Green and Barber 1981; Gregg and Purdy 2007; Jerger et al. 1988; Stroop 1935); but Simon (1969) and some others had studied a variation of visual Stroop effects with the key press response (e.g., Simon 1969, 1982).

Key press response was introduced to study the interference of direction in Simon effect as the action of key pressing in different hands would be interfered by directional cue (Simon 1969). The success of Simon's study (1969) had raised the use of key press response into wide application for Stroop experiments. Benefits from the advantages of key press response in comparing to the oral response can then be brought into the Stroop studies. As key press response can

be collected and measured by electronic devices, the measurement of key press response would be more accurate than the manual measurement (MacLeod 2005). In addition, by using electronic measurement, the experiment can exclude the presence of the experimenter (MacLeod 2005). The concrete key pressing action may also avoid faked oral response (i.e., uncertainty delayed in oral response, such as: um, uh-hah, I think, etc.) (MacLeod 2005). Using key press response can maximize the accuracy of data collection.

2.2 Purpose of Study and Research Question

Although researches in Stroop effects were common, there were only limited numbers of audiovisual Stroop studies (e.g., Cowan and Barron 1987; Peng and Prime 2014), and most of them were studying the unidirectional model of how auditory distraction interfered with visual stimulus (e.g., Cowan and Barron 1987). As such, this study intended to be an audiovisual Stroop study examining the two-directional modal of how the audio and visual components distract and interfere with each other.

Conditioned sound is an unstudied audio stimulus especially in the field of Stroop effect. As it is paired with meaning, it could be used to study the Stroop effect. Yet, until now, no examination was made on the conditioned sound. This study did used conditioned sound to study the Stroop effect.

The research question for this study is asking whether the perception process of pictures and conditioned sounds may differ from that of words in the Stroop effect based on the theory of automaticity.

There are four research hypotheses in the present study as follows:

- H1** The incongruent condition would exert greater interference than the congruent condition;
- H2** Visual distractor would exert greater interference than auditory distractor;
- H3** Vocal sound would exert greater interference than conditioned sound; and
- H4** Words would exert greater interference than pictures.

2.3 Method

2.3.1 Design

This experiment has a between-subject and within-subjects design. The experiment used 2 test conditions (congruent/incongruent) $\times$ 2 audio stimuli (spoken word/conditioned sound) $\times$ 2 visual stimuli (written word/picture) within-subjects design. In addition, the between-subject design involves two task conditions

Table 2.1 The eight combinations of condition variables, audio variables, and visual variables

Condition	Auditory stimuli	Visual stimuli	
		Word	Picture
Congruent	Word	Congruent	Congruent
		Spoken word	Spoken word
		Written word	Picture
	Conditioned sound	Congruent	Congruent
		Conditioned sound	Conditioned sound
		Written word	Picture
Incongruent	Word	Incongruent	Incongruent
		Spoken word	Spoken word
		Written word	Picture
	Conditioned sound	Incongruent	Incongruent
		Conditioned sound	Conditioned sound
		Written word	Picture

(choose answers according to the visual stimuli/choose answer according to the audio stimuli) in counterbalanced task design. In the experiment, task A referred to participants choosing answers corresponding to the audio stimuli, and task V referred to participants choosing answers corresponding to the visual stimuli. Table 2.1 shows the combinations of the above variables.

2.3.2 *Participants*

Sixty participants (30 male and 30 female participants) were recruited in this study. The recruited participants were physically healthy without any auditory, visual, or finger motion defects. They were college students aged between 17 and 24 and continuously lived in Hong Kong in the past 10 years.

2.4 **Materials**

The research materials used in task A and task V were identical. It were presented on a computer device (screen size: 27 cm high × 34 cm wide; resolution: 1280 wide × 1024 high; orientation: landscape; brightness: 100 %) with the E-PRIME software (Schneider et al. 2001). It included eight combinations of stimuli as listed in Table 2.1 above. Each combination had included eight task items. Hence, each task contained 64 task items and the whole experiment contained totally 128 task items.

Each of the 128 task items first showed a fixation page with a “+” (in the middle of the page; bolded; font: Courier New; size: 50) for 250 ms, then it would automatically change to the question page until a response was received, and finally changed to a blank page for 100 ms after the response was given in the question page. Afterwards, it would automatically change to another fixation page to start another task item. Each question page consisted of an audio stimulus (either a conditioned sound or a spoken word in a specific theme, e.g., USB insertion, pedestrian traffic light, etc.; volume: 25 in window 7; presented in 90 ± 1 Hz) which started to present together in joint presentation with the visual stimulus in white background (either a picture or a written word; in a specific theme corresponding to the audio stimuli; picture presented in the middle of the screen, resolution: 1280 wide $\times$ 1024 high; written words presented in font “新細明體,” size: 50, bolded) circled by a red circle (used to highlight the visual image) and the answer choices (either a picture or a written word as same as the visual stimulus; in a specific theme corresponding to the visual stimulus presented above; picture presented in the two sides at the bottom of the screen, resolution: 1280 wide $\times$ 1024 high; written word presented in font “新細明體,” size: 40, bolded) presented on the same screen at the same time. A letter “A” or “L” (in capital letter; bolded; font: Calibri; size: 50) was placed right above the answer choice to indicate that participant might press either the letter key “A” or “L” in the keyboard provided for answering each task item. Participants should answer as soon as possible after the answer choices were presented in each task item, which then ended the task item. The 128 task items were continuously carried out in this sequence until the whole task had been completed. The above would be repeated for every task item.

The time to complete all of these 128 task items by one participant was approximately 8 min, whereas the whole process including the training task was approximately 15 min.

2.4.1 Procedures

The experiment was divided into two parts, mainly to separate the two tasks (task A and task V) by a break so as to minimize the mixing up of the two tasks.

Before the start of the experiment, the experimental content and task instruction was first explained to the participants, then followed by a trial exercise consisted of 16 task items. Feedback showing their accuracy and response time under the trial exercise was provided after each task item for 1000 ms to increase their understanding to the operation of the experiment.

A recognition training was then presented to ensure the participants could recognize the images and conditioned sound used in the experiment later on.

After that, the experiment was started and participants were required to answer the 64 task items in the first task (either task A or task V). A 2-min break was then given, followed by the second part of the experiment.

In the second part of the experiment, the instruction for the required task (either task A or task V opposite to the first part) was first provided and then followed by a trial exercise, as those carried out in the first part. After making sure they got the correct way to do the task, the second task started.

Participants were required to finish all of the 128 task items in both parts of the experiment. The response time and accuracy of each task item were recorded immediately by the E-Prime software during the test.

2.4.2 The Control of the Extraneous Factors

Several controlling measurements used to minimize the extraneous factors were implemented in the study.

The recognition training was used to identify and train-up the less familiar stimuli (especially conditioned sound) to limit the individualized differences of conditioning effect. The restriction in participant recruitment on those continuously living in Hong Kong for the past 10 years might ensure the accessibility of the local stimuli used in the experiment.

A big red circle placed surrounding the visual stimuli was to enhance the participants both to look at the visual stimuli and listen the audio stimuli before giving their responses with a view to reduce the possibility of the attentional bias that the participants might choose not to look at the visual distractor in the audio task; while it was assumed that the participants should be distracted equally by the visual and audio distractor in both tasks in the experiment.

2.5 Results

A $2 \times 2 \times 2 \times 2$ four-way repeated-measure ANOVA was conducted with average reaction time of corrected response and sum of accuracy as the dependent variables, and tasks (visual task, audio task); congruence conditions (congruence, incongruence); visual stimuli (picture, word); together with the audio stimuli (conditioned sound, voice) as the independent variables. Table 2.2 and Table 2.3 showed the mean and standard deviation of the average reaction time and sum of accuracy, respectively, for each independent variable.

2.5.1 Average Reaction Time

There was a significant main effect for tasks [$F(1,944) = 146.39, p < .001$, partial $\eta^2 = .13$], with those doing the visual task ($M = 863.69$, $SD = 273.47$) reporting

Table 2.2 Means, standard deviations, and AVOVA for average reaction time

	<i>M</i> (SD)	ANOVA
Task		
Visual task	863.69 (273.47)	146.39***
Audio task	1105.55 (357.69)	
Congruence condition		
Congruent condition	951.31 (322.71)	11.11**
Incongruent condition	1017.94 (354.50)	
Visual stimuli		
Picture	1012.93 (368.11)	8.02**
Word	956.31 (308.09)	
Audio stimuli		
Conditioned sound	1026.71 (380.63)	17.74***
Spoken word	942.53 (289.14)	

** $p < .01$; *** $p < .001$

Table 2.3 Means, standard deviations, and ANOVA for sum of accuracy

	<i>M</i> (SD)	ANOVA
Task		
Visual task	6.78 (1.09)	9.78**
Audio task	6.58 (1.22)	
Congruence condition		
Congruent condition	7.11 (1.07)	183.63***
Incongruent condition	6.24 (1.09)	
Visual stimuli		
Picture	7.08 (1.13)	158.10***
Word	6.27 (1.05)	
Audio stimuli		
Conditioned sound	6.60 (1.21)	6.13*
Spoken word	6.75 (1.11)	

* $p < .05$; ** $p < .01$; *** $p < .001$

significantly shorter average reaction time than those doing the audio task ($M = 1105.55$, $SD = 357.69$). Hypothesis 2 was confirmed.

There was a main effect for congruence conditions [$F(1,944) = 11.11$, $p < .01$, partial $\eta^2 = .01$], with those under congruent condition ($M = 951.31$, $SD = 322.71$) reporting significantly shorter reaction time than those under incongruent condition ($M = 1017.94$, $SD = 354.50$). Hypothesis 1 was confirmed.

There was a main effect for different visual stimuli [$F(1,944) = 8.02$, $p < .01$, partial $\eta^2 = .01$], with those using words as visual stimulus ($M = 956.31$, $SD = 308.09$) reporting significantly shorter reaction time than those using pictures as visual stimulus ($M = 1012.93$, $SD = 368.11$). Hypothesis 4 was confirmed.

There was also a main effect for different audio stimuli [$F(1,944) = 17.74$, $p < .001$, partial $\eta^2 = .02$], with those using voice as audio stimulus ($M = 942.53$, $SD = 289.14$) reporting significantly shorter reaction time than those using conditioned sound ($M = 1026.71$, $SD = 380.63$). Hypothesis 3 was confirmed.

Apart from the main effects, there was a significant task and audio stimuli interaction with average reaction time [$F(1,944) = 21.49$, $p < .001$, partial $\eta^2 = .02$]. In the visual task, the responses with spoken word as distractor ($M = 867.94$, $SD = 262.58$) would have longer reaction time and that with conditioned sound as distractor ($M = 859.45$, $SD = 284.42$) would have shorter reaction time. On the other hand, in the audio task, responses toward spoken word ($M = 1017.12$, $SD = 295.68$) would have shorter reaction time and responses toward conditioned sounds ($M = 1193.98$, $SD = 391.51$) would have longer reaction time.

It was also found that there was a significant tasks $\times$ congruence conditions $\times$ audio stimuli interaction [$F(1,944) = 4.38$, $p < .05$, partial $\eta^2 = .01$].

2.5.2 Sum of Accuracy

The result showed a significant main effect for tasks [$F(1,944) = 9.78$, $p < .01$, partial $\eta^2 = .01$], with those doing the visual task ($M = 6.78$, $SD = 1.09$) reporting significantly higher accuracy than those doing the audio task ($M = 6.58$, $SD = 1.22$). Hypothesis 2 was confirmed.

There was also a significant main effect for congruence conditions [$F(1,944) = 183.63$, $p < .001$, partial $\eta^2 = .16$], with those having congruent condition ($M = 7.11$, $SD = 1.07$) reporting significantly higher accuracy than those having incongruent condition ($M = 6.24$, $SD = 1.09$). Hypothesis 1 was confirmed.

In addition, there was a significant main effect for different visual stimuli [$F(1,944) = 158.10$, $p < .001$, partial $\eta^2 = .14$], with those using pictures as visual stimulus ($M = 7.08$, $SD = 1.13$) reporting significantly higher accuracy than those using words as visual stimulus ($M = 6.27$, $SD = 1.05$). Hypothesis 4 was rejected as the result showed a reversed pattern.

There was also a main effect for different audio stimuli [$F(1,944) = 6.13$, $p < .05$, partial $\eta^2 = .01$], with those using voice as audio stimulus ($M = 6.75$, $SD = 1.11$) reporting significantly higher accuracy than those using conditioned sound as audio stimulus ($M = 6.60$, $SD = 1.21$). Hypothesis 3 was confirmed.

There was a significant task and audio stimuli interaction with sum of accuracy [$F(1,944) = 12.38$, $p < .001$, partial $\eta^2 = .01$]. In the responses in visual task, spoken word as distractor ($M = 6.74$, $SD = 1.11$) would have a lower accuracy and those with conditioned sound as distractor ($M = 6.81$, $SD = 1.08$) would have a higher accuracy; meanwhile in the audio task, responses toward spoken word ($M = 6.75$, $SD = 1.11$) would have a higher accuracy and responses toward conditioned sounds ($M = 6.60$, $SD = 1.21$) would have a lower accuracy.

2.6 Discussion

Three out of four hypotheses in the study were confirmed by the results, except hypothesis 4 which was rejected due to the showing of an opposite pattern of result for sum of accuracy.

According to the attention network model, the three sub-functions of attention, namely alerting, orienting, and executive attention have their specific function (Posner and Rothbart 2007). In this research, alerting and orienting had been kept constant in all situations, and only executive attention had been measured as conflicts brought by the different sensory information arisen in the task. Selective attention might also present in the experiment as the experimental design had instructed the participants to answer only toward one sensory stimulus, and participants might choose to focus on the attended information and ignore the unattended information.

As a cross modal Stroop task, the major discussion would be whether the visual or audio channel have exerted greater distraction on the other. The experiment had replicated the result of Peng and Prime (2014), showing that visual stimuli would exert greater distraction to the task than audio stimuli. Therefore, when participants were required to complete the audio task (to response towards what they heard), as having been distracted largely by the visual stimuli, the responses would become slower and more inaccurate.

2.6.1 Audiovisual Interaction in a Cross Modal Task

The significant main effects in congruence condition aspect had confirmed that the Stroop tasks had triggered interference effect in the experiment. The results had supported other studies under incongruent condition, which concluded that the responses would have longer average reaction time and lower sum of accuracy (e.g., McGurk and MacDonald 1976; Peng and Prime 2014; Schriefers and Meyer 1990).

It showed that the audiovisual cross modal interaction effect was present with the incongruent condition. As there was only one distractor in each condition, and that distractor was another sensory stimuli (i.e., if visual stimulus is the target stimuli, audio stimuli would be the distractor, and vice versa), the Stroop effects shown under the incongruent condition must be due to the interference brought by the interaction between the two different sensory stimuli, resulting in a longer reaction time and lower accuracy. Therefore, this study supported the results of other cross modal studies that the audiovisual interaction would affect the perception of information and result in altering the response as showed under the incongruent condition of the task (e.g., Laurienti et al. 2004; McGurk and MacDonald 1976; Schriefers and Meyer 1990).

2.6.2 Meaning and the Automaticity Effect

The research question of this study was whether the perception process of picture and conditioned sound were the same as word. With the assumed presence of automaticity effect on word, this study hypothesized that the perception process of picture and conditioned sound was different with that of word. The results showed that both written and spoken word would trigger faster response than picture or conditioned sound. According to automaticity effect, word required less attention and resources to be perceived as it was well-practiced (Stirling 1979). The result thus supported that the presence of automaticity effect had taken role in the perception of word, thus shortened the time needed to perceive the word.

One of the evidence of the experiment was the significant main effect showed under different audio stimuli with slower reaction time and lower sum of accuracy for response with conditioned sound when compared with spoken word. This might be explained by the automaticity effect of listening to word (Stirling 1979) as hypothesized at the beginning of this study. As spoken word was automatically perceived with its meaning, while conditioned sound would require more effort to perceive its meaning, the reaction time for answers with spoken word was faster. Thus, this study had supported the presence of automaticity effect on spoken word.

It was interesting to note that for the responses towards visual stimuli, written words may trigger faster response, while picture might have higher accuracy. This result had rejected the findings of similar researches that picture would give slower reaction time and lower accuracy (e.g., Cattell 1886; Fraisse 1969). This could be explained by the data-driven error of automaticity effect. Data-driven error takes place when the incoming sensory information has overridden the intended variables in an automatic action sequence (Reason 1990). Thus, the responses toward automatic processed written word may more easily be affected by the audio distractor in comparing with picture which requires more attention to be perceived. Another possible explanation is that the picture used in this study and other studies (e.g., Cattell 1886; Fraisse 1969) are different in nature. The pictures used in this study are highly accessible in our daily life and these pictures could be well-conditioned with its meaning. In opposite to the conditioned sound, people usually make judgement based on visual aid, including those picture used in this study. Therefore, participants could give very accurate response when they saw these familiar pictures. The pictures used in other visual Stroop studies are normal line drawings (e.g., Geng et al. 2014; Lupker 1979), and therefore do not show higher accuracy than written word in their studies.

By comparing visual and audio stimuli, the results showed that both picture and conditioned sound would produce longer reaction time for response, while picture would trigger higher accuracy but conditioned sound would lead to a lower accuracy when comparing with written and spoken word, respectively. This confirmed that both picture and conditioned sound were perceived differently from word, hence required different time to be processed in the perception processes. Thus, the results in this study supported that picture and conditioned sound were having

different modes of perception from that of word. Further studies can focus on the difference between picture and conditioned sound in perception and interpretation process as to compare with words.

2.6.3 Selective Attention

There are multiple different attentional models proposed to explain the mechanism of selective attention, such as the Broadbent's model, the selective filter model, and the late-filter model. All these selective attention models had proposed that the filter of blocking most irrelevant information (only personal information can pass through the filter) was placed before the short-term memory stage, and would not affect the response making (Broadbent 1958; Deutsch and Deutsch 1963; Moray 1959; Norman 1968; Wood and Cowan 1995). However, this study did not support these models but showed that even though information were coming from different sensory organs and were interpreted as different sensory information, they did interact with each other to alter the response even participants had tried to eliminate the distraction of one of them.

In fact, this research supported the attenuation model proposed by Treisman (1960) for selective attention. Please refer to Fig. 2.1 for the Treisman's attenuation model. The attenuation model proposes that all stimuli inputs, including both attended and unattended information are equally passed through the sensory register (equally received by the sensory organs). When they pass through the attenuation control, the capacity of the unattended information would be limited, while the attended information would not be limited. Finally, with the different capacity of various information they would pass through the perceptual processes and get into the short-term memory before responses were given. In this way, as the unattended information has limited capacity, the response would be more affected by the attended information (the target stimuli in this study), even though the unattended information (the distractors in this study) still has the capacity to interact with the attended information in order to alter the response. Therefore, even the executive attention is working to eliminate the irrelevant information, the conflict brought by the incongruent condition (conflict between the meaning of visual

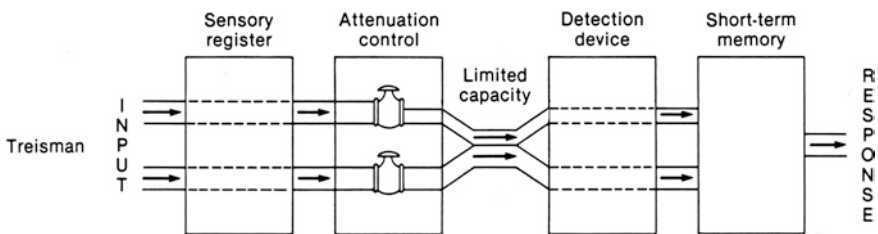


Fig. 2.1 Treisman's attenuation model

and audio stimuli) may still lead to poorer performance (longer reaction time and lower accuracy).

Therefore, this research supported the attenuation model with evidence that the capacity of unattended information had been reduced by selective attention but it still had the capacity to alter the responses triggered by the attended stimuli.

2.7 Implications

This study has addressed the interaction effect of picture and conditioned sound as comparing with written and spoken words. The results showed that spoken word would have shorter response time and higher accuracy than conditioned sound. This implies that usage of conditioned sounds as indicators in most daily instruments and tools (such as the pedestrian traffic lights in transportation system, the payment devices in public transports and stores, etc.) could be considered to be replaced by spoken words in order to increase the effectiveness of the instruments and tools to the users. This might be particularly useful for those instructions requiring high compliance or quick response. For example using spoken words in the pedestrian traffic lights can insert greater interaction effect on pedestrians and the chance of traffic accident could be greatly reduced.

Apart from using spoken word to replace conditioned sound as the automaticity effect of spoken word could provide greater interaction effect, the implication in visual aspect might be different. The results in this study showed that in the perceptual process, picture might give higher accuracy and written word might give faster response. To maximize the effectiveness of the instruments and tools, it is suggested that picture and written word can be used together as indicators to replace using only written word or picture sign separately in our daily applications. The use of combining picture and written word in instruments to improve the effectiveness of it is in fact similar to the use of spoken word (instead of conditioned sound) as mentioned in the above paragraph.

To sum up, refining instruments and tools for different aspect, such as the transportation system, medical devices, or banking transaction tools, etc., by using spoken instructions and combining picture with written word would be useful suggestion from the results of this study in order to increase the compliancy of users.

2.8 Limitations

Although every effort was employed to minimize the drawbacks of this study, there are two major limitations on this study as mentioned below.

The facilitation effect brought by the congruence of two stimuli as mentioned in some literature (e.g., Hamers and Lambert 1972; McLeod 1998; Stirling 1979) was not included in this study as limited accessible stimuli with conditioned sound

could be used to construct a neutral condition. Further study could focus on the facilitation and interference with conditioned sound to verify whether facilitation also takes place with the presence of conditioned sound.

The individualized conditioning effect may affect the current results. However, it is also the nature of conditioned sound that people seldom really pay much attention or be very familiar with the conditioned sound. Thus, this research still reflected the actual circumstances of cross modal interaction effect with conditioned sound. Further study can investigate the interaction effect with different extent of conditioning and whether conditioned sound with greater conditioning effect may override the automaticity effect of spoken word.

2.9 Conclusion

This study had introduced the cross modal interaction effect given by picture, conditioned sound and word. The findings had demonstrated that all of the stimuli, including the conditioned sound, would create interference under incongruence of visual and audio stimuli. The results proved that the automaticity effect of spoken word may trigger faster and more accurate responses than that of conditioned sound, while picture may trigger higher accuracy and written word may give faster response. Therefore, it is suggested that the daily use of conditioned sound could be replaced by spoken word and those visual signs could be modified by combining both picture and written word in our daily instruments and tools employed by various disciplines so as to increase their effectiveness.

Acknowledgments This study has obtained the ethical approval from the Research Committee of the Hong Kong Shue Yan University.

References

- Acosta, E., Jr., & Simon, J. R. (1976). The effect of irrelevant information on the stages of processing. *Journal of Motor Behavior*, 8, 181–487.
- Broadbent, D. E. (1958). *Perception and communication*. Oxford: Pergamon.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10, 649–657.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11, 63–65.
- Cohen, G., & Martin, M. (1975). Hemisphere differences in an auditory Stroop test. *Perception and Psychophysics*, 17, 79–83.
- Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception and Psychophysics*, 41(5), 393–401.
- Dalrymple-Alford, E. C. (1968). Interlingual interference in a colornaming task. *Psychonomic Science*, 10, 215–216.
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80–90.

- Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory and Cognition*, 1, 106–120.
- Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica*, 30, 96–103.
- Geng, J., Schnur, T. T., & Janssen, N. (2014). Relative speed of processing affects interference in Stroop and picture-word interference paradigms: evidence from the distractor frequency effect. *Language, Cognitive and Neuroscience*, 29(9), 1100–1114.
- Golinkoff, R. M., & Rosinski, R. R. (1976). Decoding, semantic processing, and reading comprehension skill. *Child Development*, 47, 252–258.
- Green, E. J., & Barber, P. B. (1981). An auditory stroop effect with judgements of speaker gender. *Perception and Psychophysics*, 30(5), 459–466.
- Gregg, M. K., & Purdy, K. A. (2007). Graded auditory stroop effects generated by gender words. *Perceptual and Motor Skills*, 105, 549–555.
- Hamers, J. F., & Lambert, W. E. (1972). Bilingual interdependencies in auditory perception. *Journal of Verbal Learning and Verbal Behavior*, 11, 303–310.
- Imani, A., Pourbakht, A., Akbari, M., & Kashani, M. M. (2012). Effect of sound conditioning on click auditory brainstem response threshold shifts in guinea pigs. *Audiology*, 21(3), 77.
- Jensen, A. R., & Rohwer, W. D., Jr. (1966). The Stroop color-word test: A review. *Acta Psychologica*, 25, 36–93.
- Jerger, S., Martin, R. C., & Pirozzol, F. J. (1988). A developmental study of the auditory Stroop effect. *Brain and Language*, 35, 84–104.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, 158, 405–414.
- Lidwell, W., Holden, K., & Butler, J. (2010). *Universal principles of design, revised and updated: 115 Ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design*. Beverly, MA: Rockport Publishers.
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition*, 7(6), 485–495.
- MacLeod, C. M. (2005). The Stroop task in cognitive research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 17–40). Washington, DC: American Psychological Association.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- McLeod, C. M. (1998). Training on integrated versus separated Stroop task: The progression of interference and facilitation. *Memory and Cognition*, 26, 201–211.
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2002). Multisensory visual-auditory object recognition in humans: A high density electrical mapping study. *Cerebral Cortex*, 14, 452–465.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56–60.
- Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, 75, 522–536.
- Peng, D. C., & Prime, S. L. (2014). Clarifying the crossmodal stroop effect in an auditory-visual colour naming task with words and non-words stimuli. *Journal of Vision*, 14(10), 446.
- Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58, 1–23.
- Reason, J. (1990). *Human error*. New York: Cambridge University Press.
- Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development*, 48, 643–647.
- Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, 46, 247–253.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.
- Schriefers, H., & Meyer, A. S. (1990). Experimental note: Cross-modal, visual-auditory picture-word interference. *Bulletin of the Psychonomic Society*, 28(5), 418–420.

- Settlage, P., & Harlow, H. F. (1936). Concerning the sensory pathway in the conditioned reflex. *Journal of Comparative Psychology*, 22(2), 279–282.
- Seymour, P. H. K. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, 29, 245–265.
- Simon, J. R. (1969). Reactions toward the source et stimulation. *Journal of Experimental Psychology*, 81, 174–176.
- Simon, J. R. (1982). Effect of an audio stimulus on the processing of a visual stimulus under single-and dual-tasks conditions. *Acta Psychologica*, 51, 61–73.
- Simon, J. R., Acosta, E., Jr., & Mewaldt, S. P. (1975). Effect of locus of warning tone on auditory choice reaction time. *Memory and Cognition*, 3, 167–170.
- Simon, J. R., & Berbaum, K. (1988). Effect of irrelevant information on retrieval time for relevant information. *Acta Psychologica*, 67, 33–57.
- Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The ‘Stroop effect’ vs. the ‘Simon effect’. *Acta Psychologica*, 73, 159–170.
- Simon, J. R., & Pouraghabagher, A. R. (1978). The effect of aging on the stages of processing in a choice reaction time task. *Journal of Gerontology*, 33, 553–561.
- Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audio-visual speech integration: Evidence from the speeded classification task. *Cognition*, 92, B13–B23.
- Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology*, 31, 121–132.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242–248.
- Van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech in the human brain. *Neuron*, 43, 271–282.
- Van der Molen, M. W., & Keuss, P. J. G. (1981). Response selection and the processing of auditory intensity. *Quarterly Journal of Experimental Psychology*, 33A, 177–184.
- Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one’s name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 255–260.

Applied Psychology Readings

Selected Papers from Singapore Conference on Applied
Psychology, 2016

Leung, M.-T.; Tan, L.-M. (Eds.)

2016, XIV, 248 p. 39 illus., Hardcover

ISBN: 978-981-10-2795-6