ABSTRACT

Perception is heavily influenced by attention’s limited capacity. When attention is strained between modalities and tasks, perceptual processing may be altered. This alteration can be expressed through illusory conjunctions and reaction times. Illusory conjunctions are misperceptions that result from attentional strain between multiple stimuli. From this attentional strain, a perception occurs that is an incorrect combination of features of stimuli. The current experiment examined if cross-modal illusory conjunctions between vision and audition can be induced by comparing reaction times between cross-modal control, congruent, and incongruent stimuli. Control stimuli consisted of only visual information, congruent stimuli consisted of visual and auditory information with similar features, and incongruent stimuli consisted of visual and auditory information with contradicting features. Perceptual processing was altered in a counterintuitive way previously unsupported. The current experiment’s surprising findings suggest there may be different attentional capacity limitations for different features of stimuli.

Keywords: attention, perception, cross-modal, illusory conjunctions, congruent
Introduction

Attention is the factor determining how stimuli are perceived. Attention can completely alter perception (Carrasco, 2009) resulting in inaccurate perceptions. An accurate perception is dependent on the amount of attention dedicated to stimuli. This is demonstrated in circumstances such as failure to notice something in the line of vision or mistaking features of one object for another. When inaccurate perceptions occur, it is a reflection of the attentional capacity. The attentional capacity demonstrates attention as a limited resource. There is not enough attention to disperse equally between stimuli in the environment and the mode through which they are perceived. Therefore, when attention is strained across stimuli or modalities, misperceptions can occur. Misperceptions show the attentional capacity’s profound influence over perception.

William James provides a classical explanation of attention (1890).

…taking possession in the mind, in clear and vivid form, one out of several simultaneous possible objects or trains of thought. Focalization, concentration of consciousness are its essence. It implies withdrawal from some things in order to deal effectively with others (as cited in Green, 2004).

Attention facilitates perception in occurring by maintaining an alert state in order to respond to stimuli (Posner & Fan, 2007). Working memory plays a vital role in attention facilitating perception. Working memory is a subsystem of memory that allows stimuli to be stored and used over brief periods of time (Miyake & Shah, 1999). Working memory is used to recall stimuli and events which leads to the experience of perception. A greater availability of attention directly corresponds to a greater chance stimuli will be properly stored in working memory (Cherry, 1953). Subsequently, proper storage of stimuli in working memory increases the
chances of an accurate perception. From working memory, perception is drawn. However, if the
attentional capacity is strained across stimuli or modalities, attention will not be quite as
allocated to stimuli as it would without the strain. Attentional strain can result in improper
storage of stimuli in working memory, ultimately resulting in an inaccurate perception.

Attention’s capacity is demonstrated in fixation, filtration, and attenuation of stimuli, all
of which determine perception. The attentional capacity has been compared to a spotlight
(Posner, 1980), a bottleneck (Broadbent, 1958), and an attenuating factor constraining perception
(Treisman, 1964). All three theories demonstrate that it is attention’s limited capacity that
influences perception. Perception is influenced by the attentional capacity limiting dispersal of
attention, therefore, limiting the amount of stimuli that can be thoroughly processed. When
stimuli do not receive thorough processing, they become susceptible to an inaccurate perception.
Attention’s limited capacity heavily determines how stimuli are perceived.

Broadbent’s Bottleneck Theory of attention (1958) illustrates attention’s limited capacity
as a bottleneck serving as a filter. Broadbent asserts the purpose of the bottleneck is a way to
coherently deal with constant incoming stimuli because it is impossible to allocate attention to all
stimuli (Broadbent, 1958). Attention filters out stimuli that do not fit through the perceiver’s
“bottleneck.” The bottleneck refers to characteristics of stimuli that are important to the
perceiver. Characteristics of stimuli that are not important to the perceiver become filtered out of
the bottleneck and do not reach attentional processing. Perceivers have an automatic set of
characteristics that are important to them both constantly and depending on the situation. These
important characteristics are referred to as the attentional set. One’s own name is an example of
an attentional set and is demonstrated in the Cocktail Party Phenomenon (Cherry, 1953). The
Cocktail Party Phenomenon asserts that characteristics referring to the self automatically capture
attention, such as one’s own name. Stimuli that fit the attentional set receive the most attention, whereas stimuli that are not part of the attentional set receive the least attention. The stimuli receiving the least amount of attention become filtered out of the bottleneck leading to thorough processing. When stimuli do not receive thorough processing, they have not reached working memory, and inaccurate storage can result in inaccurate perception.

Posner’s Spotlight Theory of attention (1980) also asserts attention serves as a filter. According to Posner (1980), attention is a spotlight upon a broader spectrum of area. Though stimuli encounter all area, only stimuli that fall within the circumference of the spotlight receive thorough processing. This means that out of a plethora of stimuli, only stimuli receiving attention will be perceived. It is attention, not sensation, which is the door to perception. Similar to Broadbent’s Bottleneck Theory of attention (1958), there are not enough attentional resources to encompass the entire area of stimuli, therefore, the attentional capacity serves as a filtration mechanism determining what will be perceived. Only stimuli that are filtered through the spotlight are processed thoroughly by working memory. When stimuli do not fall within the spotlight of attention, working memory does not process the information. Lack of working memory processing of stimuli may result in a perceptual phenomenon called inattentional blindness (Rock, Linnett, Grant & Mack, 1992). Inattentional blindness can result in an entirely different perception than what actually exists because of a failure to notice a characteristic of an event. Simons and Chabris (1999) demonstrated that even seemingly obvious activity can occur without being perceived due to attention’s limited capacity.

Attention’s limited capacity can result in a failure to perceive stimuli, but it can also result in an enhanced perception of stimuli. An enhanced perception of stimuli occurs when attention is allocated to stimuli (Stormer, McDonald & Hillyard, 2009). When attention is
allocated on stimuli they can appear brighter than stimuli outside of the focus of attention (Carrasco, Ling & Read, 2004). Likewise, when stimuli are stored in working memory, enhanced perception identification of stimuli is seen (Soto, Wriglesworth, Bahrami-Balani & Humphreys, 2010). Attention plays a pivotal role over how stimuli are perceived. Attention’s great influence over perception demonstrates itself as a precursor to perception.

Attention’s influence over how stimuli are perceived is demonstrated in Treisman’s Attenuation Model of attention (1964). Inaccurate perception is a result of attention’s limited capacity. Though attention’s capacity is limited, Treisman (1964) does not assert that lack of attention results in not perceiving stimuli. Instead, Treisman (1964) determined that stimuli outside the attentional set or attentional filter are still are perceived, but misperceived. This misperception is a result of stimuli not receiving focused attention. According to Treisman’s Attenuation Model of attention (1964), perception is a result of a dual-stage process. The first stage is coined the preattentive stage where all stimuli are processed to an extent. The more intense stimuli are to a perceiver, the more likely they are to reach the second stage, the focused attention stage. In the focused attention stage, stimuli are attentively focused on in order to properly store features in working memory. Proper storage in working memory results in proper recall. Stimuli that do not reach the focused attention stage remain in the preattentive stage and their features become attenuated, or weakened. This attenuation of stimuli features results in improper working memory storage. Improper working memory storage is capable of interfering with recall, which leaves recall susceptible to misperceptions. One misperception that can occur is called an illusory conjunction.

Perception is not always an accurate reflection of what physically exists. Illusory conjunctions occur when attention’s limited capacity is strained. This strain often only allows
the preattentive stage of attention to occur (Treisman, 1964). This lack of focused attention allows features of stimuli such as shape and color to be “free floating” in memory. When features of stimuli are “free floating,” they may not be able to properly integrate into a whole, accurate picture. Features of stimuli may integrate into an inaccurate, singular perception (Treisman & Gelade, 1980) called an illusory conjunction. Without the focused attention stage, features of stimuli do not always integrate into an accurate picture. This is because focused attention is considered to be the “glue” which integrates features into an accurate, whole picture. This is the basis of Treisman and Gelade’s (1980) Feature Integration Theory (FIT). When features of stimuli are not “glued” together by attention, misperceptions such as illusory conjunctions may occur.

Broadbent (1958), Posner (1980), Treisman (1964) and Treisman and Gelade’s (1980) theories all rest upon the finding that attention’s limited capacity influences how stimuli are perceived. Attention limits the amount of stimuli capable of being stored and the amount of modalities that can attend to stimuli. Attention’s limited capacity is the factor determining what, how, and if stimuli are perceived; if stimuli are perceived as they physically exist, if they are perceived at all, or if they are perceived as illusory conjunctions.

Any form of straining attention enhances vulnerability to misperceptions, such as illusory conjunctions. This strain may be an overload of stimuli, or it may be too many modalities straining to pay attention to stimuli. One circumstance of multiple modality strain resulting in illusory conjunctions is seen in the McGurk Effect (McGurk & MacDonald, 1976). When the visual component of speaking is presented simultaneously with the auditory component of sound, people frequently perceive a sound that was not an auditory stimulus presented, but instead, visually presented. This is because the features of visual and auditory stimuli did not
receive focused attention (McGurk & MacDonald, 1976). The McGurk Effect (1976) is a clear demonstration that attention’s limited capacity can be strained between multiple modalities. This is a demonstration of how vital of a role attention plays in perception. Perception is affected by attentional strain not only between multiple stimuli, but multiple modalities.

Illusory conjunctions are not always a result of attentional strain between modalities. If features of stimuli presented to multiple modalities are congruent, the attentional strain is lessened. It is apparent that attention is less strained when presented with congruent stimuli to multiple modalities because of the ability to react faster to congruent stimuli. Marks (1987) demonstrated that congruent stimuli do not strain attention as much as incongruent stimuli. Marks (1987) showed a cross-modal interaction between vision and audition by presenting a high-pitched ton simultaneously with a high-luminance light, which are congruent stimuli. Participants have significantly faster reactions to multimodal presentations when features of stimuli are congruent. Because reaction times are slower when incongruent stimuli are presented to multiple modalities, it may imply that features of incongruent stimuli require more attention and therefore are attenuated and improperly stored in memory. Improper storage may result in “free floating” stimuli and ultimately, an illusory conjunction.

The McGurk Effect (1976) shows that when attention is strained across modalities and features of stimuli, attention’s limitations may result in perception of a novel hybrid of the physical stimuli: an illusory conjunction. Marks’ (1987) experiment demonstrates that when attention is strained across modalities and presented with congruent stimuli, attention’s limitations are less strained, therefore, less susceptible to perception of an illusory conjunction. Attention’s limited capacity is an undeniable influence over perception. Ultimately, attention’s limited capacity may lead to misperceptions when attention is strained. When attention is
strained between multiple features and multiple modalities, attentional strain should result in an illusory conjunction. However, when attention is less strained by multi-modal presentation of congruent features of stimuli, perception is less likely to be distorted. A multi-modal, incongruent presentation of multiple stimuli should affect people’s accuracy in perception and take them longer to react. A multi-modal, singular presentation of stimuli should result in a higher accuracy in perception and take people a shorter amount of time to react.

Attention’s limited capacity is a profound influence over perception, and because of this, the current experiment predicts that attentional strain will result in inaccurate perceptions. The primary hypothesis is that when attention is strained between multiple modalities with incongruent stimuli, illusory conjunctions may occur. The secondary hypothesis is that congruent stimuli will receive faster reaction times than incongruent stimuli. These hypotheses are based upon Broadbent (1958), Posner (1980), Treisman (1964) and Treisman and Gelade’s (1980) findings that attention has a limited capacity that influences perception and Marks’ (1987) finding that congruency can lessen attentional strain.

Not all stimuli encountered are perceived. According to Posner (1980), stimuli that fall within the range of sensation will not be processed unless they also fall within the spotlight of attention. In order for stimuli to be thoroughly processed, not only do multiple-modalities have to be within the spotlight of attention, but also the features of stimuli. Treisman’s Attenuation Model of attention (1964) and Treisman and Gelade’s (1980) Feature Integration Theory are also supportive of the current experiment’s hypotheses because of their effects on attentional strain. Because attention is so limited, not all modalities and features of stimuli will receive the same degree of attention, therefore, some will become attenuated and susceptible to illusory conjunctions. Attention’s limited capacity may result in cross-modal illusory conjunctions
between vision and audition to be induced.

**Method**

**Participants**

Fifty-two undergraduate students at the University of Wisconsin Superior completed the study and received credit for Psychology classes. No participants disclosed having a vision or hearing deficit.

**Materials and Design**

E-Prime was used to design stimuli presented to participants. A computer screen presented participants with visual stimuli. Visual stimuli consisted of a black dot on a white screen that varied in location from upper, middle, or lower portions of the screen. Fixation points and blank screens were designed to be placed between dot screens. Headphones presented participants with auditory stimuli. Auditory stimuli varied among white noise, a high pure pitched tone (2,000Hz), a medium pure pitched tone (700Hz), and a low pure pitched tone (300Hz). A serial response box measured reaction times of participants when indicating the location of the dot on the computer screen. Three buttons on the serial response box corresponded to either the upper, middle, or lower portion of the screen. Participant accuracy and reaction time was coded on E-Prime and averaged for each combination on Microsoft Excel. Data was statistically analyzed using a within subjects one-way analysis of variance (ANOVA) through the Statistical Package for Social Sciences (SPSS).

**Procedure**

Participants were seated approximately two feet in front of a computer screen. A serial response box was placed in front of their dominant hand with labels indicating which buttons corresponded to the upper, middle, and lower portion of the screen. Participants were presented
with headphones to wear.

A black fixation point on the middle of the screen occurred for 50 milliseconds. Subsequently, a black dot on the white screen appeared, either in the upper, middle, or lower section. Simultaneously, a noise was presented on the headphones at a comfortable volume. The noise varied between white noise, a high pure pitched tone (2,000Hz), a medium pure pitched tone (700Hz), and a low pure pitched tone (300Hz). Random controlled combinations of a dot location and noise occurred 180 times. Following the 50 millisecond dot and tone combination, a white screen appeared for an infinite amount of time. Participants were instructed to decide where the previously seen dot was located by pressing a button on the serial response box during this time. Three buttons corresponded to either the upper, middle, or lower portion of the screen. After participants decided where the dot was located, the same sequence of trials continued, varying in combination of dot location and noise. Upon completion, participants were informed that the experiment was complete.

**Results**

A within subjects one-way analysis of variance (ANOVA) was conducted to test the hypothesis that cross-modal illusory conjunctions between vision and audition can be induced by measuring the relationship between the three stimulus types (control condition, congruent stimuli condition and incongruent stimuli condition). The ANOVA tested specifically to assess if there were differences in accuracy between congruent and incongruent cross-modal stimuli and to see if there were differences in reaction times between congruent and incongruent cross-modal stimuli. No significant differences were found in accuracy between conditions (F (2,49) = 1.783, MSe = 5.676)). A corresponding ANOVA revealed a main effect in reaction times (F (2,50) = 4.152, MSe = 3.123)). Planned comparisons between mean reaction times are significantly
slower for congruent stimuli than for incongruent stimuli ($t(50) = 2.434, SE = 5.02, p > .05$), significantly slower for control stimuli than congruent stimuli ($t(50) = -2.509, SE = 5.89, p > .05$), and no significant differences in the mean reaction times between control stimuli and incongruent stimuli ($t(50) = -.433, SE = 5.43$)). The means for participant accuracy and reaction times can be found in Table 1.

**Discussion**

Perceptual processing between vision and audition was altered. This altered processing was only seen in the reaction time data. Incongruent stimuli were responded to faster than congruent stimuli and control stimuli. Responses to the congruent stimuli and the control stimuli did not differ. The differences in reaction time likely reflect a difference in processing of the stimuli, otherwise all stimuli would have been responded to in a similar manner. The task was a relatively simple one, and the simplicity of the task likely resulted in the null results for the accuracy. The task was not difficult enough to strain attention enough to impair accuracy, though the actual processing was altered.

Significant differences in reaction times between conditions reinforce Broadbent (1958), Posner (1980), Treisman (1964) and Treisman and Gelade’s (1980) theories that attention has a strong influence over how stimuli are perceived. Although alteration of perceptual processing did not include cross-modal illusory conjunctions, it is important to note how theories of attention may explain this. Broadbent (1958) and Posner (1980) believe that attention serves as a filter that allows certain stimuli to reach working memory. Other stimuli are filtered out and do not reach working memory. As a consequence, the stimuli that do not reach working memory will not be perceived. Stimuli that reach working memory are part of an attentional set. The attentional set can include stimuli that are important to the perceiver. In the current task, the
attentional set would include dot location because of task demands. Noise, however, would not be included in the attentional set because the task required no noise-based decisions. Thus, because attention serves as a filter, only the dot location (attentional set) received thorough processing and resulted in high accuracy. In order to have accurate perception, all stimuli must be in working memory. This is a necessary but not sufficient condition because stimuli need to be properly integrated into working memory, not just stored. In the current experiment, noise was not processed or perceived. Thus, the weakness of these two theories (Broadbent, 1958 and Posner, 1980) is that they do not explain the differences in reaction times seen in this experiment.

Though the previous theories cannot explain the current experiment’s results, Treisman and Gelade’s (1980) Feature Integration Theory (FIT) offers a more successful perspective. It is possible that cross-modal illusory conjunctions between vision and audition were not induced because participant’s attentional capacity was not strained enough to improperly integrate features of stimuli (Treisman & Gelade, 1980). If participants were given a more attention-straining task, illusory conjunctions might be demonstrated. Due to task demands, participants consciously allocated attention on dot location resulting in high accuracy. This allocation of attention weakened attention focused on the noise. This in support of Treisman and Gelade’s (1980) FIT because all stimuli receive some degree of attention. Location was important to the task; therefore, dot location received focused attention. Features of dot location have focused attention so they are properly stored in working memory for accurate recall. Noise did not require any action for the task; consequently, noise received only pre-attentive processing. The features of the noise were not properly stored in working memory and were “free floating.” Thus, attention was strained between modalities. When working memory receives proper integration of features there is optimal processing and correct perception. The reaction time
differences in the current experiment reflect improper integration.

While the current experiment’s results can be explained by existing theories, they are contrary to some previous results. Marks (1987) found that reaction times are faster in congruent stimuli in an experiment that similarly focused on vision and audition. Different aspects of the visual stimuli were important for the task in Marks’ experiment than in the current experiment. Marks’ (1987) methodology for visual stimuli rested upon intensity of light while the current experiment utilized dot location. Marks’ (1987) experiment specifically sought to see if the pitch of a tone will result in the intensity of a light being judged with faster reaction times. It is intuitive to hypothesize that congruent information will result in faster reaction times than incongruent information, yet the results of the current experiment are contradictory to this finding. Marks (1987) hypothesized that the higher the pitch, the faster participants would react to a more intense light. A high pitch and intense light are defined as congruent information. Perhaps this suggests that perception is reliant on the specific features of attended stimuli, such as intensity or spatial location. There may be different attentional capacity limitations to different types of stimuli. Fewer features may take up less of the attentional capacity. Similarly, not all features may be equally salient. Thus, the difference between the current results and Marks’ (1987) results may just be that intensity is a more salient feature than location. The salience of these features may be determined by task demands or the inherent characteristics of the feature. If this relationship is demonstrated through all theories of attention, it may suggest that certain tasks are less likely to fall distorted to perception due to attention’s limited capacity. This hypothesis is a potential avenue for future research.

Attention is a precursor that determines how stimuli are perceived. Whether stimuli are perceived accurately or inaccurately is dependent on attention’s limited capacity. The current
experiment reveals a novel demonstration of attention’s limited capacity. Previous research has demonstrated that number of features and modalities can strain the attentional capacity. The current experiment proposes that features of stimuli may also influence the allotment of attention. An attentional capacity is more than a set amount of attention available for all stimuli. The attentional capacity is a broad term that encompasses further constraints, all of which ultimately shape perception.
### Appendix A

**Table 1**

*Participant Accuracy and Reaction Time in all Variables*

<table>
<thead>
<tr>
<th>Dot Location and Tone Combination</th>
<th>Accuracy</th>
<th>Reaction Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>93.03</td>
<td>592.01</td>
</tr>
<tr>
<td>Congruent</td>
<td>93.16</td>
<td>606.58</td>
</tr>
<tr>
<td>Incongruent</td>
<td>94.01</td>
<td>594.36</td>
</tr>
</tbody>
</table>

*Accuracy is a representation of the average percentage of trials in each condition participants were correct; reaction time is a representation of the average reaction time for dot location decisions by participants for each condition in milliseconds.*
References


