

THE EFFECTS OF STATISTICAL LEARNING AND CONGRUENCY ON THE  
DEVELOPMENT OF MULTI-MODAL OBJECTS

by

Zara Po Yee Chan

B.A. (Spec. Hons.) Glendon College, York University, 2011

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Arts

in the Program of

Psychology

Toronto, Ontario, Canada, 2013

© Zara Chan, 2013

## AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

# THE EFFECTS OF STATISTICAL LEARNING AND CONGRUENCY ON THE DEVELOPMENT OF MULTI-MODAL OBJECTS

by

Zara Po Yee Chan

For Ryerson University, Master of Arts, Psychology, 2013

## **Abstract**

The effects of statistical learning and congruency on multi-modal binding were examined. As pattern acquisition is stronger for within-object than for between-object associations, extending bias from within-object to within-modality was tested, and the statistical learning effect on between-modality learning assessed. Dyson and Ishfaq's (2008) paradigm was adapted, with frequency of within- and between-modality associations manipulated (Experiment 1), and frequency and congruency manipulated (Experiment 2). Each experiment comprised baseline (no predictive value), intra-modal (intramodal predictive value), and inter-modal (intermodal predictive value) conditions. Experiment 1 showed faster performance for within-object judgments, and fewer errors on within-object judgments, excluding the inter-modal condition. Experiment 2 replicated this, with cross-experimental analyses showing weak congruency effects. Data showed probability manipulations led mostly to interference on same-modality trials rather than facilitation on different-modality trials, suggesting while frequency of different-modality associations did not facilitate superior performance, perhaps expectancies of frequent different-modality associations weakened sensitivity to the within-modality bias.

## **Acknowledgments**

I am grateful to my supervisor, Dr. Ben Dyson, for the guidance, support, and insight that he has given me throughout my work on this project, and thank him for his mentorship and patience. I would like to extend my thanks to my thesis committee members, Dr. Meg Moulson and Dr. Todd Girard, as well as my fellow H.E.A.R. lab members, Rajwant Sandhu and Jonathan Wilbiks, for their invaluable input. I would like to thank my parents, my sister Lisa, and my friends for all of their love and support throughout the years, and for helping me develop and encouraging me to pursue my passions. Finally, I would like to thank all of my study participants, without whom the completion of this study would not be possible.

## Table of Contents

Author’s Declaration.....	(ii)
Abstract .....	(iii)
Acknowledgments.....	(iv)
Table of Contents .....	(v)
List of Tables .....	(vii)
List of Figures .....	(ix)
List of Appendices .....	(x)
Introduction.....	1
Multi-modal Binding .....	1
Statistical Learning .....	3
Visual Statistical Learning.....	6
Auditory Statistical Learning.....	8
Audio-visual Statistical Learning .....	11
Present Study .....	15
Pilot Study.....	18
Method .....	18
Preliminary Results.....	22
Preliminary Discussion .....	23
Experiment 1 .....	24
Method .....	24
Results.....	25
Discussion.....	26
Experiment 2.....	29

Method .....	33
Results.....	34
Discussion .....	35
Additional Analyses.....	36
General Discussion .....	39
Future Directions .....	45
References.....	47

## List of Tables

Table 1: Summary of Predictive Values .....	54
Table 2: Median RTs and Error Rates for All Cells in Pilot Study .....	55
Table 3: Summary of ANOVA results for RTs in Pilot Study .....	56
Table 4: Summary of ANOVA results for Error Rates in Pilot Study.....	57
Table 5: Median RTs and Error Rates for All Cells in Experiment 1.....	58
Table 6: Summary of ANOVA results for RTs in Experiment 1 .....	59
Table 7: Summary of ANOVA results for Error Rates in Experiment 1.....	60
Table 8: Transfer of Mappings .....	61
Table 9: Median RTs and Error Rates for All Cells in Experiment 2.....	62
Table 10: Summary of ANOVA results for RTs in Experiment 2 .....	63
Table 11: Summary of ANOVA results for Error Rates in Experiment 2.....	64
Table 12: Summary of ANOVA results for RTs for Cross-Experiment .....	65
Table 13: Summary of ANOVA results for Error Rates for Cross-Experiment .....	67
Table 14: Median RTs and Error Rates for Baseline Congruency in Experiment 2.....	69
Table 15: Summary of ANOVA results for RTs in Baseline Congruency in Experiment 2.....	70
Table 16: Summary of ANOVA results for Error Rates in Baseline Congruency in Experiment 2.....	71
Table 17: Median RTs and Error Rates for Baseline Congruency in Experiment 2 on Trials with Congruent-Association Prompts.....	72
Table 18: Summary of ANOVA results for RTs in Baseline Congruency on Trials with Congruent-Association Prompts.....	73

Table 19: Summary of ANOVA results for Error Rates in Baseline Congruency on Trials	
with Congruent-Association Prompts .....	74



## List of Figures

Figure 1: Depiction of Stimuli .....	75
Figure 2: Median Reaction Time for All Conditions in Pilot Study .....	76
Figure 3: Mean Error Rates All Conditions in Pilot Study .....	77
Figure 4: Median Reaction Time for All Conditions in Experiment 1 .....	78
Figure 5: Mean Error Rates for All Conditions in Experiment 1 .....	79
Figure 6: Median Reaction Time for Condition x Modality in Experiment 1 .....	80
Figure 7: Mean Error Rate for Condition x Modality in Experiment 1 .....	81
Figure 8: Median Reaction Time for Modality x Order in Experiment 1 .....	82
Figure 9: Mean Error Rate for Modality x Order in Experiment 1 .....	83
Figure 10: Median Reaction Time for All Conditions in Experiment 2 .....	84
Figure 11: Mean Error Rate for All Conditions in Experiment 2 .....	85
Figure 12: Median Reaction Time for Modality x Order in Experiment 2 .....	86
Figure 13: Mean Error Rate for Modality x Order in Experiment 2 .....	87
Figure 14: Mean Error Rate for Condition x Modality in Experiment 2 .....	88
Figure 15: Median Reaction Time for Condition x Modality x Experiment in Cross Experiment Analysis .....	89

## List of Appendices

Appendix A: Participant Consent Agreement.....	90
Appendix B: Debriefing Form .....	93

## INTRODUCTION

### Multi-modal Binding

All interactions with the world are experienced through exposure to information sensed through the various modalities of vision, audition, taction, gustation, and olfaction. Insufficient or inadequate information may be provided through an individual modality, and so, most often, the information that we gather about specific objects is multi-modal. Much of our information about the environment is sensed through vision and audition, due to the increased accessibility for more detailed information in these two modalities. Vision is undisputed as the dominant sense for humans (Calvert, Spence, & Stein, 2004), given that approximately 70% of all sensory receptors are found in the eye (Pasternak, 2005) and resources for visual processing comprise nearly half of the cerebral cortex (Sereno et al., 1995). Audition is another critical sense required for successful navigation of one's perceptual world, as it not only provides alerting information and cues for localization of objects, but also enables a level of environmental interaction that approaches that of vision (Schiffman, 2001). In addition, audition may serve as an alerting mechanism for vision (cf. Kubovy, 2001). Overall, making accurate audio-visual associations is important as it allows us to act according to our past experiences with and expectations of an object or event, thus enabling us to function more efficiently, particularly in instances where sensory information is degraded. One such situation would be using visual information to compensate for degraded auditory speech information when perceiving face-to-face speech in noisy environments (Sumby & Pollack, 1954). As we begin to associate information from one domain with information from another, multiple factors are implicated in the process of audio-visual binding, including temporal and spatial contributions (e.g., Calvert, Spence, & Stein,

2004), statistical learning (e.g., Ernst, 2007), and congruency (e.g., Molholm, Ritter, Javitt, & Foxe, 2004).

The likelihood of audio-visual binding is outlined in the *unity assumption* proposed by Welch and Warren (1980): when two or more pieces of information from separate modalities are presented, an individual must perceive them either as separate pieces of information generated by multiple events, or, associated information stemming from the same event. In terms of temporal information, a distinction between physical simultaneity and perceived simultaneity in multi-modal presentations can be made, due to the difference in the amount of time required for cortical responding to visual stimuli and auditory stimuli. Auditory stimuli are processed approximately 30 to 70 ms faster than visual stimuli at the initial stages of cortical processing (Fujisaki, Shimojo, Kashino, & Nishida, 2004), and this processing discrepancy is thought to have developed in order to provide compensation for the speed difference between light and sound transmission in air. Thus, auditory and visual signals that are physically non-simultaneous may be perceived as simultaneous if all stimulus presentations occur within a specific window of time known as the *temporal window of integration* (e.g., van Wassenhove, Grant, & Poeppel, 2007). Therefore, while two or more objects may not be presented simultaneously, differences in vision and audition with respect to cortical responding may give rise to the perception of simultaneity.

Spatial information is another factor that influences multi-modal binding, whereby multiple pieces of information relayed from a single location, holding temporal factors constant, lead to a greater likelihood of being perceived as originating from one event than do pieces of information relayed from multiple locations. Similar dissociations between physical and

perceptual localization also exist, as in the case of audition with interaural time differences, whereby a sound released from a single source will lead to differences in the time of arrival at left and right ears (Schiffman, 2001). Not only are these physical differences accommodated to assume a single source, but time differences also assist in determining the location of the source in azimuth, with a high degree of fidelity (Wallach, Newman, & Rosenweig, 1949).

Although temporal and spatial factors are both informative in influencing judgments of audio-visual binding, these factors appear to be biased towards one aspect of the audio-visual stimulus. When competition arises between visual and auditory information, the nature of the task at hand most often predicts which modality will exert greater influence over the other. For spatial tasks, visual information tends to dominate over auditory information, as demonstrated by the traditional ventriloquist illusion (Howard & Templeton, 1966). For temporal tasks, auditory information tends to dominate over visual information, as demonstrated by the auditory driving illusion (Shams, Kamitani, & Shimojo, 2000) and the temporal ventriloquism effect (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In order to control for temporal and spatial factors in the present study, both visual and auditory stimuli should be presented simultaneously and originate from the same location.

### **Statistical Learning of Associations**

In addition to temporal and spatial factors, statistical learning also influences the binding of multi-modal information. Statistical learning refers to the learning of patterns or associations in the environment, and is a process that occurs often without awareness (Kim, Seitz, Feenstra, & Shams, 2009). Statistical learning is somewhat related to classical conditioning, as both involve multiple events and are forms of associative learning. Learning by way of classical conditioning

is indicated by successful elicitation of an expected response (Dawson & Grings, 1968). The famous anecdote of Pavlov's dogs illustrates this process, as the constant pairing between the sound of a bell and the presence of food led to requiring only the sound of a bell to induce salivation. Following conditioning, responses that were previously elicited only by a related causal event are then elicited by an event paired with the related causal event, provided spatial and temporal proximity are, once again, met. In contrast, learning using statistical regularities is traditionally tested using a recognition task at test, where participants, having completed a training phase, are presented with two stimuli and must decide which stimulus is more familiar / preferable. Learning using statistical regularities is indicated by greater recognition or preference for stimuli or patterns that were presented during the training phase. Statistical learning has been empirically demonstrated in numerous studies, including experiments involving infants learning auditory (Saffran, Aslin, & Newport, 1996) and visual artificial grammar (Kirkham, Slemmer, & Johnson, 2002), the learning of visual (Fiser & Aslin, 2002) and auditory artificial grammar in adults using non-linguistic (Leboe & Mondor, 2007) and linguistic stimuli (Saffran, Johnson, & Aslin, 1999), and studies using audiovisual stimuli (Conway & Christiansen, 2006; van den Bos, Christiansen, & Misyak, 2012); Walk & Conway, 2011).

The study of language acquisition in infants and young children is of great interest in statistical learning research, due to its representing of one of the most important instances of learning in early life. Language acquisition may also inform multi-modal learning research as it involves exposure to the multi-faceted visual and auditory components for speech perception, and the conditions under which infants expect them to occur together. Statistical learning studies have traditionally focused on the use of artificial grammar paradigms that employ a training phase and a test phase. During the training phase, participants are presented with strings of

letters, structured such that the order of letters in each string is governed by various grammatical rules. During the test phase, participants are presented with novel and familiar strings that are either grammatically correct or incorrect, and asked to judge the correctness of each string. In a study with 8-month-old infants, Saffran, Aslin, and Newport (1996) examined statistical learning by looking at the impact of exposure to nonsensical words on the ability to discriminate between words and non-words. During the training phase, infants were presented with a continuous speech stream for 2 minutes, which comprised four nonsensical words of three syllables each that were repeated in random order and provided no temporal cues for word boundaries. For example, the speech stream might consist of segments such as *bidakupadotigolabubidaku*, where the first nonsensical word *bidaku* is followed by *padoti*, *golabu*, and *bidaku* again. Transitional probabilities differed between syllable pairs, thereby helping to establish within-word and between-word boundaries. For example, in the sample speech stream segment, *bida* would constitute a within-word syllable pair with a transitional probability of 1.0 (*bi* is followed by *da* 100% of the time), while *kupa* would constitute a between-word syllable pair with a transitional probability of 0.33 (*ku* is followed by *pa*, *go* or *bi* 33% of the time), giving within-word syllable pairs an advantage in predictability. To be clear, within-word syllable pairs always had a transitional probability of 1.0 since the first syllable in each pair was followed consistently by the next syllable in the word. Between-word syllable pairs had a transitional probability of 0.33 due to the fact that each syllable that concluded a word was followed with the first syllable from any of the other 3 words. Test trials involved listening to repetitions of familiar and novel words and non-words, during which infants were monitored for their listening preference. Results showed that infants learned to discriminate between words and non-words by showing longer listening times for non-words in comparison to words, indicating that they had learned the words with

only a short training phase of 2 minutes. As the rapid learning ability of infants could have been limited to audition, research was needed to investigate whether this learning mechanism would be modality-specific or modality-general. Therefore, Kirkham, Slemmer, and Johnson (2002) conducted a study similar in design to the Saffrin et al. (1996) study, using visual stimuli. Three groups of infants (aged 2 months, 5 months, or 8 months) were familiarized with sequences of discrete visual stimuli, wherein the order of stimuli was constructed according to a transitional probability. Stimuli consisted of 6 differently-coloured and discrete shapes. At test, infants were presented with alternations of familiar and novel patterns, and monitored for their looking preference. Results found that infants generally recognized the familiar patterns of visual stimuli, which was demonstrated by longer looking times for novel patterns. All infants exhibited this preference, regardless of age. The occurrence of such rapid familiarization and recognition in infants points to the impact of both visual and auditory statistical learning in the acquisition of knowledge early in life. Away from the complexities of language acquisition, statistical learning can be examined using more basic stimulus attributes and less complex grammars. Such studies may utilize stimuli comprising letters, colours, and shapes in order to provide multiple dimensions in the visual domain, and different timbres, dynamics, and varying frequencies of pitch in the auditory domain. It will now be necessary to consider how both forms of statistical learning have been studied in adults, and as well, how these two forms might collaborate or compete with one another.

### *Visual statistical learning*

As an initial insight into visual statistical learning, Fiser and Aslin (2002) presented participants with 4 sequences of 3 shapes in the familiarization phase, with a shape emerging and



disappearing from behind a vertical occluder before the appearance of the next shape. This was done with the intention that no spatial or temporal boundary cues were provided that would inform the participants of grouping among shapes. During test, participants were presented with a pair of triplets for each trial, wherein one triplet was from one of the original 4 sequences and the other was a novel triplet sequence, and asked to indicate which triplet was more familiar. Results showed that participants accurately discriminated the original triplets from the novel triplets, indicating that learning must have occurred from the association of any two shapes that repeatedly appeared together in succession. In addition to statistical learning occurring at a perceptual level, Brady and Oliva (2008) studied statistical learning at a conceptual level. During familiarization in Experiment 1, participants were presented with a stream of 12 images, which was comprised of 4 sequences of triplets. Stimuli consisted of various scenes depicting a bathroom, bedroom, bridge, building, coast, field, forest, kitchen, living room, mountain, street, and waterfall. Participants were given the task of indicating each time an image was repeated within the sequence in order to prevent any explicit awareness of the sequence structure. During test, participants were presented with a pair of triplet scenes for each trial, wherein the order of one triplet was familiar and the other was novel, and asked to indicate which triplet was more familiar. Results showed that participants correctly identified previously presented triplet sequences as being familiar above chance. Through further experimentation, Brady and Oliva (2008) also found that perceptual statistical learning of image order was transferred to learning of the order for categories that represented these images. Naïve participants were familiarized with the same perceptual stimuli used in previous experiments, and test items in this further experimentation consisted of words for the category of the images rather than the actual images themselves. For example, a training item would show a triplet image sequence of a mountain,

kitchen, and street, while the test item would show the words “mountain”, “kitchen”, and “street”. Participants were successful in correctly identifying word triplets that had been previously presented in image form, which demonstrated that statistical learning could occur on the basis of conceptual as well as perceptual information, and transfer over from perceptual to conceptual representations.

### *Auditory statistical learning*

In the same way that basic shapes can be used for visual statistical learning, basic sound attributes have been employed to study auditory statistical learning. Using low and high pitches presented from low or high speakers, Leboe and Mondor (2007, Experiment 1) conducted a study using a nonverbal auditory Stroop paradigm. Combinations of pairings between pitch height (low and high) and location (low speaker and high speaker) were presented with equal frequency to examine whether or not a Stroop effect was present. Participants were required to identify the pitch identity on half of the trials and the location identity for the remaining half. Results showed a Stroop effect, where responses were faster on trials for which both pitch height and location were congruent. The frequency of pairings was then manipulated in a follow-up Experiment 2 to examine the effect of statistical learning on congruency known as the item-specific congruency effect (Jacoby, Lindsay, & Hessels, 2003). Pitch and location identities were congruent for 80% of trials for one pitch and congruent for 20% of trials for the other pitch, creating 4 possible combinations of trials with congruent or incongruent pairings having low or high probabilities. For example, the high speaker presenting a high pitch 80% of the time would derive the high-probability congruent condition, the high speaker presenting a low pitch 20% of the time the low-probability incongruent condition, the low speaker presenting a low pitch 20%

of the time the low-probability congruent condition, and the low speaker presenting a high pitch 80% of the time the high-probability incongruent condition. Pitch height was counterbalanced across participants, as a participant would have heard a high pitch on 80% of trials and a low pitch on 20% of trials, while another participant would have heard a low pitch on 80% of trials and a high pitch on 20% of trials. The manipulation involving congruency and probabilities allowed the overall proportion of having half of the trials as congruent and the other half as incongruent to be maintained from Experiment 1; however, in Experiment 2, the irrelevant feature (location for pitch judgments, and pitch for location judgments) now had predictive or non-predictive value. Results showed that participants responded faster on congruent trials when the required response had been presented with a higher probability, and faster on incongruent trials when the required response had been presented with a higher probability. These results showed that statistical learning was able to enhance responding even for incongruent trials, for which performance is usually worse.

In another auditory statistical learning study, tonal sequences were also used in a sound stream with identical grammar to that used in Saffran et al., (1996) (Saffran, Johnson, & Aslin, 1999; Experiment 1). Each nonsensical word (e.g., *bupada*) was translated into tones by pairing a tone to each syllable in the triplet (e.g., DFE). This designation of syllables to tones yielded one “language”, and a second “language” was created by assigning a different tone to each syllable of the nonsensical words. The use of each language was counterbalanced during the training phase and the test phase involved both languages to control for any biases. The training phase consisted of exposure to a continuous tone stream for 7 minutes that was played 3 times, with short breaks in between the repeats of the tone stream. Results showed that participants performed significantly above chance level in the recognition test, which demonstrated that

statistical learning occurs when exposed to both linguistic and non-linguistic sounds. In addition, Tillmann and McAdams (2004) found that statistical regularities could be learned in auditory stimuli, even when some acoustical attributes of the sequences varied. The stimuli in this experiment comprised the tone E flat (311 Hz) played in 13 different timbres: French horn, trumpet, trombone, harp, vibraphone, striano (hybrid of strings and piano), harpsichord, English horn, bassoon, clarinet, vibrone (hybrid of vibraphone and trombone), guitar, and guitarinet (hybrid of guitar and clarinet). Three sets of sequences were constructed using differences in the distance between timbres (e.g., the timbre of a French horn is farther from that of a guitar than that of a trombone, given that the French horn and the trombone are both members of the brass family, while the guitar is a percussive string instrument). Three sets of sequences were constructed: S1 comprised small distances between timbres within triplets and large distances between triplets (supportive / congruent), S2 comprised large distances between timbres within triplets and small distances between triplets (contradictory / incongruent), and S3 comprised equidistant timbres within and between triplets (neutral). It was proposed that having perceptually weaker triplet boundaries (i.e., greater difference in distance between timbres within triplets in comparison to difference between triplets) would minimize the effect of statistical learning. There were two groups of participants: a control group and a learning group. The control group received no training and completed a testing phase wherein pairs of triplets were presented, and participants were instructed to choose which triplet in the pair sounded like it would be categorized as a unit in a longer sequence. The learning group underwent a training phase that familiarized them with a timbre sequence consisting of S1, S2, and S3 triplets, which was similar to previous artificial grammar designs. At test, participants were asked to discriminate between two pairs of triplets to identify which triplet had been previously presented.

Results showed that participants in the learning group were sensitive to the statistical regularities presented in the training phase, despite a triplet-grouping bias provided by the structure of timbral differences within and between triplets, such as in S2. This triplet-grouping bias is based on timbral differences that are smaller between adjacent timbres between triplets than are timbral differences between adjacent timbres within triplets. For example, one such sequence would involve a triplet consisting of bassoon, trumpet, and guitar timbres (A1, A2, A3), followed by a triplet consisting of guitarinet, French horn, and striano timbres (B1, B2, B3); here, the timbral difference from A1 to A2 is greater than the timbral difference from A3 and B1, meaning that the grouping of triplets would be categorized as contradictory / incongruent. Due to the closer timbral relationship between the outer timbres of the two triplets, this would lead to an inherent bias in forming triplets if no manipulation of frequency association were present. Overall, planned contrasts showed that performance was greater for congruent sequences over neutral sequences and greater for neutral sequences over incongruent sequences. These findings demonstrated that perceptual (dis)similarities between stimuli may still play an important role in forming associations, along with statistical regularities (see forthcoming discussion on congruency; current Experiment 2).

### *Audio-visual statistical learning*

In addition to the literature on statistical learning in separate modalities, there have been a handful of studies examining multi-modal interactions in statistical learning. A study by Conway and Christansen (2006) examined the extent to which artificial grammars could be learned in adult participants if two distinct grammars were presented concurrently in three experiments: in two modalities, in two dimensions in a single modality, and along one dimension in a single

modality. This particular design was employed to investigate how resources would be allocated within and between modalities when multiple streams are presented. Visual and auditory stimuli consisting of colours and tones were used in Experiment 1 to produce two grammars in two separate modalities (visual grammar and auditory grammar presented together), colours and shapes along with tones and non-words in Experiment 2 with two distinguishable grammars in a single modality (colour grammar and shape grammar presented together, and tone grammar and non-word grammar presented together), and two different sets of shapes in Experiment 3 with two similar grammars in a single modality (shape 1 grammar and shape 2 grammar presented together). It was noted that only visual stimuli were used in Experiment 3. Results showed that learning occurred for both grammars when grammars were presented simultaneously in separate modalities (Experiment 1) or in two dimensions uni-modally (Experiment 2). In Experiment 1, wherein a visual grammar and an auditory grammar were presented concurrently, correct responses were made for 63.5% of the visual grammar judgments (either colour grammar or shape grammar) and 70.5% for the auditory grammar judgments (either tone grammar or non-word grammar). In Experiment 2, wherein two grammars having distinct dimensions (e.g., colour versus shape) in the same modality were presented, correct responses were made for 59.5% for both of the colour and shape grammars, and 68.5% and 60.0% for the tone and non-word grammars, respectively. In comparison, for Experiment 3, wherein two grammars sharing one dimension (e.g., two sets of shapes with no discernible differences between sets) in the same modality were presented, correct responding was at 60.0% for one set, and 56.0% for the second set, the latter of which was not significant from responding at chance. This demonstrated that when grammars were presented in one dimension uni-modally, only one of the grammars could be learned, due to the strong perceptual similarities between the two grammars. These results

suggest that learning of the statistical structure was stimulus-specific rather than abstract, and, if stimulus attributes of two grammars were distinguishable enough, the grammars could be learned simultaneously; multiple grammars within the same sensory dimension, however, could not be learned concurrently, as this caused some difficulty for participants.

Findings from another study suggest that the multi-modal learning of artificial grammars is further complicated when grammars consist of items from both modalities, as a cost associated with different-modality responding relative to same-modality responding is shown (Walk & Conway, 2011). In this study, participants were familiarized with a multi-modal grammar consisting of three auditory stimuli and three visual stimuli. Each stimulus, whether visual or auditory, was paired with a visual stimulus and an auditory stimulus, such that the first stimulus had a .5 probability of preceding the visual stimulus and .5 probability of preceding the auditory stimulus. For example, A1 would precede only V1 or A2, while V1 would precede only V2 or A3, with the structure of this artificial grammar creating within-modality and between-modality associations. Results showed that while participants were able to identify when test items were grammatically correct or incorrect, this was limited to within-modality associations only. Violations in cross-modal associations were not recognized above chance level, suggesting that multi-modal stimuli may not be readily integrated during statistical learning. However, an audio-visual study demonstrated that responding for visual stimuli was facilitated when participants unknowingly made associations between pairs of auditory and visual stimuli that were frequently presented together. Unknown to the participants, the to-be-ignored auditory stimuli were contextual cues for spatial information, which provided information about the visual stimuli (Kawahara, 2007). Similarly, a study examining the effects of input from vision or audition on statistical learning in the other modality found that auditory input enhanced learning of visual

sequences when the auditory input was presented with the same statistical structure as the visual input (Robinson & Sloutsky, 2007). In addition, a study by van den Bos, Christiansen, and Misyak (2012) involving 3 experiments showed sensitivity to the type of non-adjacent dependency (either deterministic [1] or probabilistic [0.5]) in learning to discriminate between grammatically-correct and –incorrect sequences of non-word triplets. Experiment 1 showed that exposure to deterministic dependencies with auditory stimuli led to higher discriminability versus exposure to probabilistic dependencies. When visual cues were added in Experiment 2, such that arbitrary associations were made between each category of non-word (initial, middle, and final word in a triplet) and a visual feature (colour, shape, and continuous uni-directional movement), learning of both deterministic and probabilistic dependencies was successful, as it was, as well, in Experiment 3, wherein phonological cues were integrated into the auditory stimuli. The results from these studies show that statistical learning can occur cross-modally, although the degree to which statistical learning extends cross-modally is not yet clear.

Therefore, three effects are evident in the multi-modal statistical learning literature. First, it is relatively easy to maintain multiple grammars or associations if they are separated across modalities (Conway & Christiansen, 2006). Second, irrelevant information presented (e.g., input from another source) or irrelevant information provided within the target information (e.g., the location from which a stimulus is presented) influences learning (Kawahara, 2007; Robinson & Sloutsky, 2007). Third, it is relatively difficult to establish a single grammar or pattern of associations across modalities (Walk & Conway, 2011). Taken together, the results of the various aforementioned studies support the finding that the strengthening of associations between dimensional stimulus attributes and stimuli promotes binding, although this effect may be less effective for binding of multi-modal objects due to a strong within-modality bias.



Continuous pairings of stimulus attributes lead to an increased likelihood that the presentation of one stimulus in a pair will evoke quicker and more accurate responses for its complementary stimulus. While most studies show a robust effect for within-modality learning, the effect for between-modality learning is not as evident. However, given that pairing stimuli together strengthens the associations between these stimuli, it is expected that continued exposure should promote a multi-modal learning effect. Providing a temporally synchronous and spatially coincident presentation of the paired stimuli should also promote audiovisual integration, and subsequently, learning; therefore, the proposed experiments pair stimuli from separate modalities by presenting them simultaneously as a multi-modal object, rather than presenting the stimuli in succession as in the traditional statistical learning paradigm. The role of statistical regularities is a main interest for both experiments in the present study, as pairings of stimuli are manipulated to promote association between pairs, and as per van den Bos et al. (2012), deterministic associations should promote binding relative to probabilistic associations.

### ***Present study***

To examine within- and between-stimulus associations at the level of individual multi-modal stimuli, it is possible to adopt the paradigm used by Dyson and Ishfaq (2008), with this study itself adapting a task used by Duncan (1984). Applying the original Duncan (1984) paradigm to the auditory domain, Dyson & Ishfaq (2008) developed two distinct auditory stimuli: a tone with pitch and FM modulation (modulation of the frequency to induce a vibrato), and a noise with bandpass and intensity modulation. These two sounds were played simultaneously for each trial, and thus, as each sound had two distinguishable properties, each trial provided four pieces of information. Participants were required to respond to two of these

properties. Half of the trials asked about both properties of the same sound (within-object) and half of the trials asked about one property from each of the concurrently-presented sounds (between-object). The authors hypothesized that if auditory memory could be organized according to the object from which it is derived, performance on the second response for within-object trials should be better than for between-object trials. In other words, there should be privileged access to other information regarding an item in memory once that item has initially been interrogated, and this should be reflected in faster second responding for within-object trials relative to between-object trials. Although statistical regularities were not manipulated in the study, results showed that participants exhibited a within-object effect, as reaction time (RT) data showed an advantage for second responses for within-object trials. Given that this study was conducted in an auditory setting, how might these results inform a similar study design using multi-modal visual and auditory stimuli? Information may be grouped according to the object from which it originates, thus, it is plausible that information is likely grouped according to the modality from which it originates. Consistent with the within-modality bias (Walk & Conway, 2011), the bias for same-object responding in a uni-modal setting would be expected to be analogous to a bias for same-modality responding in a multi-modal setting. However, as statistical learning has been shown to increase the formation of associations between dimensions, it is hypothesized that having deterministic associations between audio-visual pairs should also promote multi-modal binding, although any effects between modalities are likely to be smaller than effects that are within-modality.

The present experiment examined the effects of statistical learning on the development of associations for same-modality objects and different-modality objects. The design was adapted from a paradigm used in Dyson and Ishfaq (2008) in order to study associations across vision

and audition, and was supported by the design used in Walk and Conway (2011). These effects were studied using initially non-congruent stimulus attributes, which are presented in Figure 1. Visual stimulus attributes consisted of square and diamond shapes (V1) in 2D and 3D depths (V2), and auditory stimulus attributes consisted of low and high pitches (A1) with or without the presence of a warble (A2; analogous to a quick vibrato). No pre-existing associations have been found between any of these stimulus attribute dimensions in the literature.

The experiment comprised 3 conditions: baseline, intra-modal, and inter-modal (see Table 1). The baseline condition presented an equal number of pairings for all combinations of stimulus attributes. Here, it was expected that second responding for same-modality objects will show enhanced performance, which constitutes decreased RTs and / or decreased error rates for judgments of the second stimulus attribute in a pair when both attributes originate from the same modality, compared to when the attributes originate from different modalities.

The intra-modal condition paired each value of a stimulus dimension with a value of the other stimulus dimension from the same modality in within-modality associations to give a deterministic probability of 1.0 [e.g., A1 value 1 (low pitch) was always paired with A2 value 1 (no warble), A1 value 2 (high pitch) was always paired with A2 value 2 (warble), etc.]. This was similar to the within-modality association pairing in Walk and Conway (2011). Between-modality associations in this condition were probabilistic, meaning there was no predictive value for cross-modal associations, given that a visual attribute did not have a consistent pairing with any one auditory attribute, and vice versa. In the intra-modal condition, it was expected that the within-modality bias exhibited in the baseline condition will be increased due to the presence of deterministic probabilities for within-modality associations.

The inter-modal condition paired each value of a stimulus dimension with a value of the other stimulus dimension from the opposite modality in between-modality associations to give a deterministic probability of 1.0 [e.g., A1 value 1 (low pitch) was always paired with V1 value 1 (square), A1 value 2 (high pitch) was always paired with V1 value 2 (diamond), etc.]. This was similar to the between-modality association pairing in Walk and Conway (2011). Within-modality associations in this condition were probabilistic and did not have any predictive value for within-modality associations. In this condition, it was expected that the within-modality bias will be decreased by speeding different-modality responding, given the presence of cross-modal deterministic probabilities.

The pilot experiment reported below examined statistical learning and associations using non-congruent stimulus attributes. Manipulations in all 3 conditions are identical to those in Experiment 1.

## **Pilot Study**

### **Method**

**Participants.** A sample of 12 undergraduate students (4 male, 8 female) participated in the study. Two participants were excluded from analyses due to high error rates (error rates of over 25%) and replaced. The mean age of the participants was 21.0 years ( $SD = 3.21$ ) and all were right-handed. Participants were recruited through the Ryerson University Research Participant Pool and received one credit towards their course credit for their participation. The study was approved for testing by the Research Ethics Board of Ryerson University. Participants gave informed consent following an explanation and trial run of the experiment, and received

debriefing on the purpose of the experiment following completion of the study (see appendices for consent and debrief forms).

**Stimuli and apparatus.** Auditory stimuli consisted of 500 ms high (523 Hz) and low (294 Hz) tones, which were with or without frequency modulation (modulation frequency 20 Hz, deviation frequency 20 Hz) to create the presence or absence of a warble, and were created using SoundEdit 16 (MacroMedia). Linear on-set and off-set ramps of 5 ms were applied. Sounds were presented binaurally from free-field speakers (Harman/Kardon) positioned on either side of a computer monitor that was viewed approximately 57 cm away, to best simulate spatial coincidence between auditory and visual stimuli (Calvert et al., 2004). Visual stimuli consisted of squares and diamonds presented in either 2-dimensional or 3-dimensional form, and were created using Powerpoint (Microsoft). Squares were 3.3cm<sup>2</sup>, and diamonds were created by rotating the squares 90° clockwise. Visual stimuli were presented in the centre of a black screen. The presentation of stimuli was controlled by PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993) and responses were recorded using a PsyScope Button Box.

**Design.** Experimental blocks of 128 trials each were developed using combinations of visual stimulus shape (V1; square, diamond) and depth (V2; 2D, 3D), and auditory stimulus pitch (A1; low, high) and warble presence (A2; no, yes). Prompts displayed the to-be-judged stimulus attribute in the centre between the two possible values. Prompts were “SQUARE[shape]DIAMOND” and “2D[depth]3D” for visual stimuli, and, “LOW[pitch]HIGH” and “NO[warble]YES” for auditory stimuli. For each of the 128 trials, two prompts were presented. Each stimulus attribute (shape, depth, pitch, and warble) was presented as the first prompt for an equal number of trials, yielding four sets of 32 trials. For each stimulus attribute, it

was also ensured that both values (e.g., shape; square, diamond) were presented an equal number of times (16 trials), meaning correct responses for the first prompt corresponded to the left button for half of the trials, and to the right button for the other half. For same-modality trials, responses for the second prompt were based on the only other stimulus attribute belonging to the stimulus interrogated at the first prompt (e.g., shape followed by depth or vice versa in the case of visual responding, and warble followed by pitch or vice versa in the case of auditory responding). For different-modality trials, responses for the second prompt were based on one of the attributes from the alternate modality (e.g., shape followed by pitch, warble followed by depth, etc.). By implementing these manipulations, a probability inequality arose with respect to between-modality combinations outnumbering within-modality combinations. Therefore, to maintain equal probability of associations, the pairings of stimulus attributes were modeled after a study by Dyson (unpublished), wherein the number of combinations for both same-modality and different-modality pairings were equal by associating each stimulus attribute with its within-modality attribute (e.g., shape and depth) and only one of the two possible between-modality attributes (e.g., shape and pitch, or shape and warble).

For first prompt responding, both values for each stimulus attribute were presented an equal number of times (16 trials). On the basis of this initial design, three conditions were developed: baseline, within-modality, and between-modality. All conditions differed in terms of the strengthening or weakening of associations between stimulus dimensions, which are shown in Table 1. In the baseline condition, all possible combinations of stimuli were orthogonal (as per Dyson & Ishfaq, 2008), meaning that there was no predictive value for both same- and different-modality associations. In the intra-modal condition, there was predictive value for within-modality associations and no predictive value for between-modality associations, (e.g., the value

of A1 would predict the value of A2 for same-modality trials but not the value of V1 for different-modality trials), whereas for the inter-modal condition, there was predictive value for between-modality associations and no predictive value for within-modality associations (e.g., the value of A1 would predict the value of V1 for different-modality trials but not the value of A2 for same-modality trials).

While predictive value would lead to advantages in RT and error rate, a potential confound dealing with the assignment of response buttons to particular stimulus attributes could lead to an inflation of the aforementioned advantages. Specifically, advantages in responding for the second response in each pair of responses could also be attributed to response repetition. For example, in shape (V1) judgments, the response buttons are left and right for square and diamond, respectively, and in pitch (A1) judgments, the left and right buttons for low and high, respectively. If shape had predictive value for pitch judgments (V1 predicts A1, as in the between-modality condition), wherein squares were associated with low and diamonds were associated with high, the button used for the second response is the same as the button for the first response. This response button repetition could inflate any advantages resulting from actual associations between stimulus dimensions. In order to eliminate this potential confound, two versions of each condition were created such that associations elicited response repetition (correct responding for both responses mapped onto the same button) or response change (correct responding for both responses mapped onto different buttons). Thus, in one version, squares would be associated with low pitch and diamonds associated with high pitch (response repetition), and in the other version, squares would be associated with high pitch and diamonds associated with low pitch (response change). Therefore, the effects of dimensional associations could be evaluated independently of response repetition and change. The order of conditions

(baseline, intra-modal, and inter-modal), between-modality associations (e.g., A1 paired either V1 or V2), and association between response buttons for first and second responses (response repetition vs. response change) were counterbalanced across participants. Trial order was randomized in both the practice and experimental blocks.

## **Procedure**

Participants were given a tutorial on the different stimuli to ensure adequate discrimination between stimulus attributes before completing a practice block consisting of 12 trials taken at random from the baseline condition. Participants then completed one block of 128 trials for each of the three conditions (baseline, within-modality, and between-modality). Each trial began with a fixation cross for 500 ms before a visual stimulus and auditory stimulus were presented simultaneously for 500 ms. Next, following the presentation of a blank screen for 500 ms, the first of two prompts regarding stimulus judgments appeared and remained onscreen until the participant made a response. Responses were made by pressing the button (left or right) that corresponded to the stimulus attribute previously seen or heard. After each response, there was a checking period for 50 ms before participants were given feedback for 450 ms in the form of a green cross for a correct response and a red cross for an incorrect response. The process was repeated for the second prompt.

## **Preliminary Results**

Median RT and mean error rates for each cell are presented in Table 2. Values from the ANOVA table for RT and error rate data are presented in Tables 3 and 4, respectively. Mean RTs and error rates for each of the three conditions (baseline, intra-modality, and inter-modality)



are presented in Figures 2 and 3. A three-way repeated-measures ANOVA (2 x 2 x 3) was conducted on RT and error rate data with the type of modality judgment (same-modality or different-modality), response order (first or second), and condition (baseline, intra-modality, or inter-modality) as within-subjects factors. Greenhouse-Geiser corrections were used as the assumption of sphericity had been violated. Although these are preliminary results, the data follow the results of Dyson & Ishfaq (2008) and extend these findings from within audition to across vision and audition. There was a marginal effect of modality,  $F(1, 11) = 4.725, p = .052, \eta^2 = .300$ , showing responses were slightly faster on same-modality trials (827 ms) than on different-modality trials (868 ms), and a main effect of order,  $F(1, 11) = 11.864, p < .01, \eta^2 = .519$ , showing that participants responded faster for second responses (802 ms) than for first responses (893 ms). A marginal interaction was revealed between the type of modality judgment and response order,  $F(1, 11) = 4.739, p = .052, \eta^2 = .301$ , indicating that second responses were slightly faster for same-modality trials relative to different-modality trials. A 3-way interaction between condition, modality judgment, and response order was not found with the current data. Error data showed an effect of response order,  $F(1, 11) = 8.175, p < .05, \eta^2 = .426$ , indicating that there were more errors to be committed on second responses across all conditions.

### **Preliminary Discussion**

In the pilot experiment, there was a marginal interaction between type of modality judgment and response order, which supports the notion that accessing pieces of information from a single object / modality may be faster than accessing pieces of information from multiple objects / modalities (Dyson & Ishfaq, 2008). One concern in the pilot experiment was the joint observation of faster RT and higher error rates for second responding (see also Dyson & Ishfaq,

2008). Although this does not compromise the marginal two-way interaction trend found between response order and type of modality judgment, it would be ideal to try to eliminate any speed-error trade-off in subsequent experiments. One possible locus for this effect was identified in the duration of the delay between the offset of stimuli and the presentation of the first prompt. The delay, set at 500 ms, may have been too short, and thus, potentially caused participants difficulty in responding as some processing of the previous stimuli may still have been taking place during the time allotted for responding to the first prompt (Dyson, unpublished; Dyson & Ishfaq, 2008). The delay was increased to 1000 ms in Experiment 1, to help improve the level of consolidation and processing for the stimuli presentation, and allow participants to be more ready to respond, which would likely result in faster RT for first responses. This increase in delay duration at 1000 ms was expected also to help allay any potential speed-accuracy trade-off, as lower error rates for first responding would be acceptable if RTs were equivalent with second responding.

## **EXPERIMENT 1**

### **Method**

24 undergraduate students (19 female) participated in the study; mean age was 20.9 years ( $SD = 5.3$ ) and 20 were right-handed. Two additional participants were excluded for error rates exceeding 30% in any conditions. The recruitment process to participation, stimuli and apparatus used, and design were identical to those of the pilot study. The procedure was also identical to that of the pilot study, other than the small change in duration of the blank screen exposure after stimulus presentation from 500 ms to 1000 ms to help improve the level of stimulus presentation processing and consolidation.

## Results

Median RTs and error rates for each of the three conditions (baseline, intra-modal, and inter-modal) are presented in Table 5. Values from the ANOVA table for RT and error rate data are presented in Tables 6 and 7, respectively. A three-way repeated-measures ANOVA (2 x 2 x 3) was conducted on RT and error rate data with the type of modality judgment (same- or different-modality), response order (first or second), and condition (baseline, intra-modal, and inter-modal) as within-subjects factors. For RT data, there were main effects of type of modality judgment,  $F(1, 23) = 17.070, p < .001, \eta p^2 = .426$ , and response order,  $F(1, 23) = 38.341, p < .001, \eta p^2 = .625$ . An interaction between type of modality judgment and response order,  $F(1, 23) = 28.728, p < .001, \eta p^2 = .555$ , was revealed, indicating that second responses were faster for same-modality trials (747 ms) relative to different-modality trials (871 ms; Tukey's HSD,  $p < .05$ ). These results confirmed again that Dyson and Ishfaq's (2008) findings were extended from within audition to across vision and audition. There was also an interaction between condition and modality,  $F(2, 46) = 8.831, p < .001, \eta p^2 = .277$ , as the manipulation of association strength between particular stimulus attributes in conditions led to greater differences in RT between responses on same-modality trials and different-modality trials. A Tukey's HSD, ( $p < .05$ ), revealed that differences were present between baseline and inter-modal conditions for responding on different-modality trials, as responses in the baseline condition (907 ms) were slower than responses in the inter-modal condition (856 ms) (see Figure 6).

Error rate data showed main effects of type of modality judgment,  $F(1, 23) = 22.564, p < .001, \eta p^2 = .495$ , and response order,  $F(1, 23) = 45.071, p < .001, \eta p^2 = .662$ . A three-way interaction between condition, response order, and type of modality judgment was revealed,  $F(2,$

46) = 6.611,  $p < .01$ ,  $\eta^2 = .223$ , indicating that less errors were committed on second responses for same-modality trials in relation to different-modality trials, except in the inter-modal condition. A Tukey's HSD, ( $p < .05$ ), showed no differences between second responding on different-modality trials across conditions (15.82%, 15.26%, and 13.24% for baseline, intra-modal, and inter-modal conditions), while a difference between second responding on same-modality trials occurred between the intra-modal (7.47%) and inter-modal (13.59%) conditions (see Figure 5). This demonstrated that error rates for different-modality trials remained consistent while error rates for same-modality trials changed as a function of condition. Results also show that the possibility of a speed-error trade-off has been reduced, given that first responses in Experiment 1 were magnitudinally slower and more errorful than first responses in the pilot study.

## **Discussion**

Data showed faster RT for second responding on same-modality trials, providing further evidence that accessing pieces of information from a single object / modality is faster than accessing pieces of information from multiple objects / modalities (Dyson & Ishfaq, 2008). Response speed was also influenced by the manipulation of association strength between particular stimulus attributes. The difference in manipulation of probabilities across conditions demonstrated that statistical learning influenced the learning of between-modality associations in the inter-modal condition. Additionally, the three-way interaction in the error rate data demonstrated that statistical learning influenced the within-modality bias as error rates for second responses on same-modality trials in the inter-modal condition was significantly different from those in the baseline and intra-modal. Interestingly, the modulation of the within-modality

bias was not due to facilitation on different-modality trials, but rather from interference on same-modality trials, running contrary to Duncan (1984) and Dyson (unpublished). This finding showed there was a lack of improvement in performance, given that the frequent co-occurrence of within-modality associations did not improve accuracy on second responses for same-modality trials in the intra-modal condition, and the increased difficulty in binding cross-modally was not eased by the frequent co-occurrence of cross-modal associations in the inter-modal condition. In the inter-modal condition, wherein associations were present for only between-modality stimulus attributes, frequent cross-modal associations not only failed to benefit performance, but also worsened performance for second same responding. Perhaps there was some constraint in responding that did not allow the learning of same-modality and different-modality associations to translate into improved performance. It is possible that probability manipulations are insufficient to influence between-modality responses; mere exposure to patterns in stimulus attribute associations may not have been enough to improve performance. As results showed in Walk and Conway (2011), participants were able to recognize violations in grammar for within-modality associations, but not for between-modality associations. The perceptual stimuli were created along a single visual dimension (shape) and a single auditory dimension (pitch). Perhaps it was easier to recall within-modality associations as each stimulus attribute was more related to the other attributes in its dimension, given that the attributes shared similar perceptual features, such as colour and the presence of edges. There was no relationship between the attributes in the visual and auditory dimensions, as although shape and pitch do have congruency (Marks, 1987), the visual stimuli were not distinct in angularity to provide high-angular and low-smooth associations. As such, it could be possible that the use of deterministic probabilities to form associations between between-modality stimulus attributes was inadequate

due to the randomness of assigning associative attributes for stimulus dimensions (although see van den Bos, Christiansen, & Misyak, 2012). As error rate data showed only an effect of interference on same-modality trials and not an effect of facilitation on different-modality trials in the inter-modal condition, perhaps the integration of related dimensions would enhance the ability of statistical learning to help form associations. While it may be beneficial to use non-arbitrary stimulus attribute associations, given the effect of congruency, it is noteworthy to mention that increased similarity amongst dimensions may not necessarily lead to improved statistical learning. In Conway and Christiansen's (2006) study, presenting two different grammars in the same dimension (i.e., shape) led to difficulties in recognizing grammatically correct sequences from grammatically incorrect sequences. The high level of similarity between stimulus attributes may have led to greater difficulty in keeping the two grammars separate, and subsequently, impairment in the processing of each grammar. However, in the present study, given that the associations are not created in temporal succession, but rather, in temporal synchrony, similarity between dimensions between attributes would not be of great concern. Given the lower rate of success in showing between-modality learning using deterministic probabilities in Experiment 1, the effect of congruency with the use of synaesthetically congruent audiovisual pairings was considered in promoting the learning of associations between modalities in Experiment 2. It was hypothesized that inherent cross-modal associations and manipulated association strength would further reduce the same- modality bias shown in Experiment 1 and increase performance for different-modality trials.

## EXPERIMENT 2

In addition to manipulating the association strength between stimulus attributes, the effect of congruency on multi-modal binding was examined. Congruency effects have been examined in uni-modal and multi-modal perceptual studies. Uni-modal stimuli can be classified as being congruent or incongruent based on the defining characteristics of the stimulus. Examples include having a high pitch being presented from a high location as a congruent stimulus (Dyson, 2010), and seeing the word “red” in red colour as a congruent stimulus (Stroop, 1935). Conversely, multi-modal stimuli are classified as being congruent or incongruent based the relatedness between the stimuli in the multiple modalities. Examples include the high pitch of an auditory stimulus presented with the visually-presented word “high” as being congruent (Melara & Marks, 1990), hearing a dog bark and seeing a picture of a cat as being incongruent (Yuval-Greenberg & Deouell, 2009), and hearing the word “red” while seeing a blue visual target as incongruent (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). Associations have been found between multiple attributes in vision and audition such as size and brightness (Walker & Walker, 2012), timbre, pitch, and loudness (Melara & Marks, 1990; Dyson & Quinlan, 2010), loudness and brightness (Marks, 1987), loudness and size (Gallace & Spence, 2006; Smith & Sera, 1992; Wilbiks & Dyson, in review), pitch and shape (Marks, 1987), pitch and brightness (Marks, 1987; Melara, 1989; Martino & Marks, 1999b), colour and pitch (Melara, 1989), lightness and vibrotactile frequency (Martino & Marks, 1999a), pitch and size (Gallace & Spence, 2006; Mondloch & Maurer, 2004), pitch and angularity (Marks, 1987), and pitch and vertical location, size, and spatial frequency (Dyson, 2010; Evans & Treisman, 2010). In vision, Walker and Walker (2012) demonstrated that the inherent relationship between the dimensions of size and brightness was such that objects that were brighter were associated with a smaller

size. Spence and Deroy (2012) argue that these perceptual relationships were formed due to an association occurring in the environment, as smaller objects are more likely to be found closer to the light source, which, in most natural environments, is from above. In audition, Melara and Marks (1990) found relationships between three attributes: pitch and loudness were correlated such that higher pitch was associated with louder volume and lower pitch with softer volume, timbre and loudness were correlated such that hollow and soft sounds were associated while “twangy” and loud sounds were associated, and timbre and pitch were correlated such that hollow and higher sounds were associated while “twangy” and lower sounds were associated. These findings demonstrate that congruency exerts considerable influence on learning and subsequent responses, and that processing of perceptual characteristics of objects in a modality is influenced by perceptual input from other modalities.

Within the realm of congruency, a distinction can also be made between perceptual correspondences and semantic congruency (Spence, 2011), the former of which is the non-arbitrary associations between physical stimulus attributes, and the latter the associations that are made based on identity or meaning. As inherent associations between different multi-modal attributes exist, comparisons have been drawn between crossmodal correspondences and synaesthesia, which is a condition wherein the perception of a stimulus attribute in one modality creates the perception of an attribute in another modality despite the lack of stimulation in the second modality (Martino & Marks, 2001; Spence, 2011). Although it is an issue not entirely resolved, some researchers currently accept the view that synaesthesia is characterized by strong crossmodal associations, and that normal associations between multi-modal stimulus attributes exist on a single continuum (Sagiv & Ward, 2006; Ward, Huckstep, & Tsakanikos, 2006). Three classes of crossmodal correspondences are described by Spence (2011): structural, statistical, and



semantic. Structural correspondence represents a shared basis of encoding magnitude between dimensions with magnitudinally based properties (i.e., loudness and brightness), such as the rate of neural firing. This correspondence requires little or no training, evidenced by 20-30 day old infants showing knowledge of associations between loudness and brightness (Lewkowicz & Turkewitz, 1980). Statistical correspondence refers to the learning of naturally occurring associations between dimensions (i.e., loudness and size). Only the physical properties of objects are considered in judgments of crossmodal correspondence, such as size, shape, and colour in visual stimuli, and loudness, pitch, and timbre in auditory stimuli. Lastly, semantic correspondence represents a shared concept between two dimensions (i.e., pitch and elevation) and is learned through the use of language and common terms. Stimuli having semantic correspondence can be abstract and require mental representations of an idea or pairing. The meaning and physical properties of objects are both considered in semantic congruency judgments. Visual and auditory presentations of a word as written and spoken language are examples of semantic congruency, as both forms of the word have the same meaning. Similarly, a picture of a baby crying and a sound of the cries also provide semantic congruency, as well as statistical correspondence, which is based on the frequency of co-occurrence in the environment (Spence, 2011). Regarding the level at which congruency plays a role, studies involving the use of congruent and incongruent stimuli show that congruency effects consist of higher accuracy and faster RT, while incongruency effects consist of lower accuracy and slower RT (Baier, Kleinschmidt, & Müller, 2006; Kim, Seitz, & Shams, 2008; Marks, Ben-Artzi, & Lakatos, 2003; Sandhu & Dyson, 2012; Shams & Seitz, 2008). It is clear that the presence of two pieces of congruent information should allow the brain to complete processing more effectively in a shorter length of time compared to two pieces of information in competition with one another, as

the positively correlated information provided in congruency leads to the same conclusion (or response output) while the incongruent information requires further processing in order to resolve the assignment of stimulus to response. This enhanced responding that arises from the presentation of the same piece of information in two dimensions is referred to as a *redundancy gain* (Melara, 1989).

Visual dimensions and auditory dimensions compiling two individual sets of cross-modal congruencies (A1 associated with V1, A2 associated with V2) were used in the second experiment of the present study to investigate the combined effects of congruency and statistical learning of associations within and between modalities. The purpose of using congruent associations between modalities was to examine the degree to which congruency contributes to statistical learning to overcome between-modality grammar issues (van den Bos et al., 2012) and how the level of association between modalities may be improved to approach that of association within modalities.

Experiment 2 sought to establish whether the combined effects of statistical learning and cross-modal congruency can lead to a between-modality bias and the manipulation of different-modality judgments, in contrast to the data in Experiment 1. To achieve the congruency effect, visual stimulus attributes consisted of round and wavy circles in white and grey colours, and auditory stimulus attributes consisted of low and high pitches with or without the presence of a warble (see Figure 1). Pitch and lightness have been shown to correspond (high pitch and white, low pitch and dark) (Mondloch & Maurer, 2004) and timbre and shape as well (pure tone and round shape, complex tone and more angular shape; Hossain, 2011). Congruent relationships present only for stimulus dimensions between modalities (i.e., shape and warble, colour and

pitch) were required to examine the influence cross-modal associations may exert on statistical learning. In Experiment 2, the combination of statistical learning and congruency was hypothesized to enhance responding for between-modality associations. In the baseline condition, second responding is expected to be comparable for both same- and different-modality trials, considering that despite having congruent pairings exist only between modalities, a same-modality bias should provide a facilitating, albeit weaker, effect. In the intra-modal condition, second responding is expected to be enhanced for same-modality trials, with a more salient within-modality bias here in comparison to the baseline condition due to the added effect of statistical learning between A1 and A2, and, V1 and V2. In the inter-modal condition, second responding is expected to be enhanced for different-modality trials due to the added effect of cross-modal correspondences between A1 and V1, and A2 and V2.

The design for Experiment 2 was identical to that of Experiment 1, apart from having visual stimuli simply changed from shapes (square and diamond) in varying depth (2D and 3D) to different shapes (circle and cloud) in varying colour (white and grey) to provide congruency with the auditory stimuli for cross-modal associations only. Table 8 illustrates the transfer of attribute and response mappings from Experiment 1 to Experiment 2.

## **Method**

24 undergraduate students (18 female) participated in the study; mean age was 21.9 years (SD = 7.3) and 22 were right-handed. 2 additional participants were excluded for error rates exceeding 30% in any condition. The recruitment process to participation was identical to that of the pilot study.

Presentation of stimuli and the auditory stimuli used remained the same from Experiment 1. Visual stimuli consisted of circle and cloud shapes in white and grey, and were created using Powerpoint (Microsoft). The diameter of the shapes was 3.3cm<sup>2</sup> and shapes were presented in the centre of a black screen. The design was identical to that of Experiment 1, with the replacement of square and diamond shapes with circle and cloud shapes (V1; circle, cloud), and 2D and 3D shapes with white and grey shapes (V2; white, grey). Auditory stimulus attributes remained pitch (A1; low, high) and warble presence (A2; no, yes). Pitch and colour were associated (high pitch-white, low pitch-grey) and warble and shape were associated (no warble-round circle, warble-wavy circle). Prompts were “CIRCLE[shape]CLOUD” and “WHITE[colour]GREY” for visual stimuli, and, “LOW[pitch]HIGH” and “NO[warble]YES” for auditory stimuli. The procedure was identical to that of Experiment 1.

## Results

Median RT and mean error rates for each cell are presented in Table 9. Values from the ANOVA table for RT and error rate data are presented in Tables 10 and 11, respectively. Mean RTs and error rates for each of the three conditions (baseline, intra-modal, and inter-modal) are presented in Figures 10 and 11, respectively. A three-way repeated-measures ANOVA (2 x 2 x 3) was conducted on RT and error rate data with the type of modality judgment (same- or different-modality), response order (first or second), and condition (baseline, intra-modal, and inter-modal) as within-subjects factors. Greenhouse-Geiser corrections were used where the assumption of sphericity had been violated. RT data showed main effects of type of modality judgment,  $F(1, 23) = 45.828, 20.279, p < .001, \eta p^2 = .666$ , and response order,  $F(1, 23) = 20.279, p < .001, \eta p^2 = .469$ , were found. Results further confirmed the extension of Dyson and Ishfaq's

(2008) findings, which were revealed in a RT interaction between type of modality judgment and response order,  $F(1, 23) = 32.583, p < .001, \eta p^2 = .586$ , indicating that second responses were faster for same-modality trials relative to different-modality trials. A Tukey's HSD, ( $p < .05$ ), revealed significant differences between responses for second same-modality trials (799 ms) and second different-modality trials (965 ms).

Error rate data showed main effects of type of modality judgment,  $F(1, 23) = 7.346, p < .05, \eta p^2 = .242$ , and response order,  $F(1, 23) = 29.845, p < .001, \eta p^2 = .565$ . A three-way interaction between condition, response order, and type of modality judgment,  $F(2, 46) = 4.297, p < .05, \eta p^2 = .157$ , was revealed, indicating that more errors were committed on second responses for same-modality trials in relation to different-modality trials in the inter-modal condition. A Tukey's HSD, ( $p < .05$ ), showed that a difference was present between second same-modality trials in the intra-modality condition (6.32%) and the inter-modality condition (11.09%). Similar to Experiment 1, error rates on second different-modality trials were not significantly different across conditions.

## **Discussion**

The data show faster RT for second responding on same-modality trials, providing further evidence that accessing pieces of information from a single object / modality is faster than accessing pieces of information from multiple objects / modalities (Dyson & Ishfaq, 2008). The three-way interaction for error rates demonstrated that statistical learning influenced the within-modality bias as error rates for second responses on same-modality trials in the inter-modal condition were significantly different (larger) from those in the intra-modal condition. Again, confirming the findings from Experiment 1, the modulation of the within-modality bias from the

error rate data was not due to facilitation on different-modality trials, but rather from interference on same-modality trials.

### **ADDITIONAL ANALYSES**

It was hypothesized that the performance on different-modality trials would be superior in the inter-modal condition of Experiment 2 compared to that of Experiment 1, exhibiting gains from using synaesthetically congruent stimuli (Mondloch & Maurer, 2004; Hossain, 2011). Therefore, a four-way repeated-measures ANOVA ( $2 \times 2 \times 2 \times 3$ ) was conducted on RT and error rate data, with type of modality judgment (same- or different-modality), response order (first or second), and condition (baseline, intra-modal, and inter-modal) as within-subjects factors, and the experiment (using non-arbitrary stimulus associations or synaesthetically congruent stimulus associations) as a between-subjects factor. Values from the ANOVA table for RT and error rate data are presented in Tables 12 and 13, respectively. Greenhouse-Geiser corrections were used where the assumption of sphericity had been violated. RT data showed an interaction between type of modality judgment and experiment,  $F(1, 46) = 4.243, p < .05, \eta p^2 = .084$ , and an interaction between condition, type of modality judgment, and experiment,  $F(2, 92) = 5.614, p < .01, \eta p^2 = .109$ . Although Tukey's HSD tests revealed there were no meaningful differences as a function of experiment, the interaction between condition and type of modality that was found in Experiment 1 was not present in Experiment 2,. Error rate data showed no significant differences as a function of experiment, signifying there was no overall effect of congruency.

Concerns regarding the failure of the congruency manipulation were further explored by examining differences across experiments. The strength of the congruency manipulation was also

investigated as the manipulation of congruency in this experiment was not as strong as was originally thought; cross-modally congruent pairings in the inter-modal condition were not present for both associations (i.e., colour and pitch, but not colour and warble, and, shape and warble, but not shape and pitch) as these pairings were counterbalanced across participants. A measurement of congruency in Experiment 2 was taken by comparing the performance for same-modality and different-modality trials in the baseline condition based on the number of between-modality congruent pairings on a given trial: 0, meaning no congruent associations were present (i.e., grey circle and high pitch with warble); 1, meaning one congruent association was present (i.e., grey circle and low pitch with warble); and 2, meaning both congruent associations were present (i.e., grey circle and low pitch without warble). As the combination of stimulus attributes contributed to a larger number of trials having one congruent pairing when compared to the number of trials having both congruent pairings or none, the number of trials was equated by randomly selecting 16 trials for each set of congruent pairing trials. Median RT and mean error rates for each cell are presented in Table 14. A two-way repeated-measures ANOVA (2 x 3) was conducted on RT and error rate data, with type of modality judgment (same- or different-modality) and congruency (no congruent associations, one congruent association, two congruent associations) as within-subjects factors. Values from the ANOVA table for RT and error rate data are presented in Tables 15 and 16, respectively. Greenhouse-Geiser corrections were used where the assumption of sphericity had been violated. RT data showed a main effect of type of modality,  $F(1, 23) = 11.437, p < .01, \eta p^2 = .332$ , showing that responses on same-modality trials (881 ms) were faster than on different-modality trials (964 ms). An interaction between type of modality and congruency was revealed,  $F(2, 46) = 4.095, p < .05, \eta p^2 = .151$ , although a Tukey's HSD ( $p < .05$ ) yielded no meaningful results, although the results suggest that for trials with no

congruent pairings, responses on between-modality trials were slower than on same-modality trials. Error rate data showed a main effect of congruency,  $F(2, 46) = 8.830, p = .001, \eta p^2 = .277$ . Interestingly, less errors were committed on trials with one congruent pairing (5.01%) in comparison to trials with no congruent pairings (9.05%) and two congruent pairings (9.31%). Perhaps it is the case that having only one congruent pairing meant that there was no distraction or less of a distraction from the non-congruent pairing (e.g., seeing a white circle and hearing a high FM sound gives cross-modal congruency for colour-pitch but not shape-warble) on responding; having two congruent pairings with only one pairing being relevant for responding create some distraction in responding, as the strength of association between congruent pairings is stronger than that of non-congruent pairings. Additionally, there is a possibility of a speed-error tradeoff, as, magnitudinally, faster responses and more errors were recorded on trials with no congruent pairings or both congruent pairings.

Given that the results from this congruency analysis involves responses on trials wherein participants were exposed to different levels of congruent pairings for presented stimuli but not necessarily congruent dimensions in the response prompts, a second analysis on congruency was conducted using only trials with prompts interrogating the attributes involved in the congruent pairing(s). Median RT and mean error rates for each cell are presented in Table 17. A two-way repeated-measures ANOVA (2 x 3) examining the effect of congruency was conducted, with trials having either no congruent pairings, one congruent pairing, or two congruent pairings, and having prompts interrogating the dimensions involved in congruent pairings (shape and warble, or colour and pitch). Values from the ANOVA table for RT and error rate data are presented in Tables 18 and 19, respectively. Greenhouse-Geiser corrections were used where the assumption of sphericity had been violated. RT data showed a marginal main effect of type of modality,  $F(2,$



22) = 11.00,  $p = .052$ ,  $\eta^2 = .300$ , showing that responses on same-modality trials (897 ms) were faster than on different-modality trials (993 ms). Error rate data showed a main effect of congruency,  $F(2, 22) = 3.548$ ,  $p < .05$ ,  $\eta^2 = .244$ . Again, less errors were committed on trials with one congruent pairing (5.99%) in comparison to trials with no congruent pairings (9.90%) and two congruent pairings (10.42%).

## GENERAL DISCUSSION

The present study investigated the effect of statistical learning (Experiment 1) and the joint effects of statistical learning and congruency (Experiment 2) on the development of multi-modal objects. Traditional statistical learning experiments have tended towards the use of a paradigm involving stimuli sequences with transitional probabilities both within and between sequences (i.e., Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). These paradigms examine the learning of transitional probabilities that were temporally separated, given that the presentation of stimuli was in succession, but do not address learning of transitional probabilities that could have occurred from simultaneous presentations. As the present study investigated multi-modal object formation, it was ideal to present visual and auditory stimuli together, rather than separately (e.g., Calvert, Spence, & Stein, 2004). The current study adapted the paradigm used in Dyson and Ishfaq (2008) for use with audiovisual stimuli. Both experiments were comprised of three conditions, each of which involved differing manipulations of within-modality and between-modality associations in order to gauge the effectiveness of statistical learning. In the baseline condition, neither same-modality or different-modality associations had predictive value as no mappings between stimulus attributes were consistent; here, a stimulus attribute had equal probability of pairing with any other stimulus

attribute. In the intra-modal condition, same-modality associations were given predictive value, such that a value in one dimension (e.g., A1: high pitch) was mapped onto a value in the other dimension in the same modality (e.g., A2: warble) to give a deterministic probability of 1.0. Similarly, in the inter-modal condition, different-modality associations were mapped such that particular pairs of values, each consisting of a value in an auditory dimension and a value in a visual dimension, were given a deterministic probability of 1.0. Predictive value was given to within-modality associations in the intra-modal condition in order to strengthen the within-modality bias, while predictive value was given to between-modality associations in the inter-modal condition to reduce within-modality bias.

In Experiment 1, RTs were shown to be faster for second responses when the second prompt inquired about a feature from the same modality as the first feature. This was in line with the results of Dyson and Ishfaq (2008) and extended the results from multiple auditory objects to single multi-modal objects. In addition, overall, participants were faster at responding to second prompts and on same-modality trials. This supported the view that memory is object-based; in other words, multiple features of singular object are more easily remembered than the same number of features on multiple objects as the recollection of a feature from one object allowed quicker access to other features of that object (Olson & Jiang, 2002). Fewer errors were committed on same-object responses as well, with the exception of the inter-modal condition, wherein error rates for second responses on both same-modality and different-modality trials were equally as high. This was an interesting result to see, for two reasons. First, the manipulation of having high transitional probabilities for between-modality associations failed to lower error rates for different-modality trials, despite the predictive nature of these pairings. It had been hypothesized that the probability manipulation would reduce the number of errors,

given that numerous studies had shown high transitional probabilities had influenced learning (e.g., Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996; Tillmann & McAdams, 2004). Instead, the data showed that errors were not reduced as a result of probability manipulation, and rather, that the only change in error was an increase for same-modality trials in the inter-modal condition. The mechanism behind the increase in difficulty is unclear, although, it is thought that the interference on same-modality trials could have been caused by the overall lack of associations for within-modality attributes in the inter-modal condition, coupled with the high probability manipulations for between-modality associations. Perhaps the difficulty in associating a visual attribute with an auditory attribute, and vice versa, was too great to be overcome with probability manipulations, meaning that the binding between audiovisual features could have been inherently insufficient. It was unexpected not only that the manipulation of having deterministic probabilities for between-modality associations failed to reduce within-modality bias, but that it also increased the difficulty for accurate responding on same-modality trials. In addition, perhaps there was some sort of constraint on the level of optimized performance for both same-modality and different-modality trials, given that the frequent co-occurrence of within-modality associations did not improve accuracy on same-modality trials in the intra-modal condition, and the increased difficulty in binding cross-modally was not eased by the frequent co-occurrence of cross-modal associations in the inter-modal condition. It is possible that probability manipulations could not improve accuracy for different-modality responses in the inter-modal condition because performance was already at its best for different-modality trials. In addition, the exposure to frequent cross-modally congruent pairings coupled with probability manipulations favouring cross-modal pairings in the intermodal condition may have led to an expectancy of stimulus presentations to have consistent predictive

value for cross-modal associations. This expectation could interfere with responses for same-modality trials, given that the discontinuation of a pattern of frequent associations between different-modality attributes should affect error rates.

In Experiment 2, results corroborated the findings from Experiment 1, as again, RTs were faster on second responses when both prompts inquired about the same modality. Similar to Experiment 1, fewer errors were committed on same-object responses with the exception of the inter-modal condition; however, the joint effect of statistical learning and congruency in this condition may have differentiated the error rates of second responding between same-modality and different-modality trials, as error rates for second responses on same-modality trials were magnitudinally higher than those for second responses on different-modality trials. Findings from Experiment 2 were similar to those of Experiment 1, as shown by the lack of significant results from the cross-experimental analyses, suggesting that congruency was not able to play a role in enhancing the effect of statistical learning on promoting multi-modal binding. A concern regarding the strength of the congruency manipulation was that the transfer of mappings from the stimulus attributes in Experiment 1 to those in Experiment 2 did not ensure that all congruent cross-modal associations would be strengthened to the same degree through statistical learning. As the experimental design required counterbalancing for the various non-congruent combinations of stimulus attributes in Experiment 1, replacing each of the shape-depth stimuli with one of the shape-colour stimuli meant that the probability manipulations embedded in the design now did not, in Experiment 2, provide all congruent cross-modal associations. This led to a weaker effect of congruency, given that the joint contribution of statistical learning and congruency did not necessarily mean statistical learning with congruent associations throughout. Although preference has been shown for different cross-modal associations [pitch and colour

(Mondloch & Maurer, 2004) and shape and warble (Hossain, 2011), this contribution of congruency may still be insufficient to effectively bolster the development of multi-modal objects, confirming that establishing cross-modal associations can be rather difficult (Conway & Christiansen, 2006; Walk & Conway, 2011). Although there was an effect of congruency in the error rate data from the congruency analysis, there was no benefit due to probability manipulations, neither for second responding on same-modality trials in the intra-modal condition nor for second responding on different-modality trials in the inter-modal condition. This is in contrast to most other statistical learning studies, as the effect of statistical learning in the present study was not shown by facilitation, but rather, interference. It could be argued that this interference does not exactly exemplify “learning”, given that the evidence of this learning was worsened performance; however, as statistical learning refers to the acquisition of pattern knowledge and the implicit awareness of associations, learning did take place because an effect due to probability manipulations was present. Thus, regardless of the nature of the outcome, participants were sensitive to changes in probability for the co-occurrence of specific attribute associations, and this pattern acquisition was reflected in their performance. As acquiring these patterns with the use of cross-modally congruent pairings in the perceptual stimulus dimensions did not produce an added or modulated effect of statistical learning, it may be the case that congruency has a salient effect when using tasks that examine familiarity / preference. Perhaps it is more difficult to produce congruency effects for perceptual cross-modal associations involving tasks that require responses for judgments other than familiarity / preference, whereas semantic cross-modal associations may be easier to establish in such task, given that semantic associations are likely more salient and familiar (e.g., Leboe & Mondor, 2007). In addition, acquiring statistical properties of associations might have been easier in the intra-modal condition versus in

the inter-modal condition due to the within-object / within-modality bias. The tendency to group attributes by the object of their origin would have been enhanced from the subsequent practice of recollecting both attributes from the same modality.

Some additional considerations, in terms of experimental design, concerned a reversal of the mappings of each stimulus attribute value to correspond to either the left and right buttons on the Button Box. This would eliminate any potential response congruency effects, which refers to enhanced responding based on the response mapping and not the stimulus attributes themselves (c.f., *decisional level*; Melara & Marks, 1990). For example, the response mapping of low pitch to the left button might be considered congruent while a mapping of low pitch to the right button might be considered incongruent since this horizontal mapping is similar to that of a piano, where lower and higher tones are produced by playing keys on the left and right sides of the keyboard, respectively (e.g., Gevers & Lammertyn, 2005). However, since response mappings are currently divided into response change and response repetition across participants when greater-than-random associations take place, this would lead to a rather large increase in the number of participants needed in order to fully counterbalance over all conditions; therefore, another option would be to include both response change and response repetition in each condition for each participant. By doing so, in addition to the elimination of potential response congruency effects, this option would also allow the analysis of any potential interaction between type of response mapping and congruency (i.e., *stimulus-response compatibility effect*), and subsequently, any potential interaction between type of response mapping and time / number of trials needed to show statistical learning effects. However, having ever-changing response mappings for each value of all stimulus attributes to left and right buttons would produce slower RTs and higher error rates overall, due to the increase in information presented to and processed

by the participants before a response could be made. This, in turn, could be countered by recording responses vocally, although this would be fairly laborious and data would be more prone to human error in scoring for accuracy and calculations of RT. Therefore, the present design is thought to be sufficient and best suited to examine the effects at hand.

## **FUTURE DIRECTIONS**

As the use of congruent pairings of perceptual stimuli did not bolster the effect of statistical learning, it would be useful, firstly, to consider using congruent pairings of semantic stimuli, given the increased saliency of the semantic associations in comparison to perceptual associations. The paradigm could remain as is, with the stimuli changed in Experiment 2 from white and grey circles and clouds, to stimuli with semantic reflections of the auditory stimuli (low and high pitch with or without warble) such as circles and clouds presented in higher or lower space. Pairing low and high pitches played from low and high speakers could also provide stronger congruent pairings (Leboe & Mondor, 2007; Sandhu & Dyson, 2013). Secondly, an additional consideration would be temporally successive presentation, given that much of the statistical learning literature employs the artificial grammar learning paradigm (Conway & Christiansen, 2006; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Leboe & Mondor, 2007; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, & Aslin, 1999; van den Bos, Christiansen, & Misyak, 2012). Presenting stimuli one after another, with visual information before auditory information, or vice versa, would reduce the potential of temporal contributions to multi-modal binding. However, this may give participants added time to consolidate the features of the stimuli being presented as they would be attending to two features at a time, instead of four. It could be argued that performance on trials interrogating features from the first

presentation would be at a disadvantage, though, given that the time elapsed from presentation to response would be greater than for trials interrogating features from the second presentation. Piloting a study to determine the effect of this temporal presentation and the type of stimuli to be used would be beneficial in seeking ways in which multi-modal binding could be enhanced. Lastly, as our task did not involve familiarity and recognition phases consistent with the traditional statistical learning paradigm, it might be worthwhile to consider the intra-modal and inter-modal conditions as potential training blocks, and presenting participants with novel and familiar pairings on trials for test blocks.



## References

- Baier, B., Kleinschmidt, A., & Müller, N. G. (2006). Cross-modal processing in early visual and auditory cortices depend on the statistical relation of multisensory information. *Journal of Neuroscience*, 26, 12260-12265.
- Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, 19, 678-685.
- Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processing*. Cambridge, MA: MIT Press.
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities. *Psychological Science*, 17, 905-912.
- Daza, M. T., Ortells, J. J., & Fox, E. (2002). Perception without awareness: Further evidence from the Stroop task. *Perception & Psychophysics*, 64, 1316-1324.
- Dawson, M. E., & Grings, W. M. (1968). Comparison of classical condition and relational learning. *Journal of Experimental Psychology*, 76, 227-231.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501-517.
- Dyson, B. J. (2010). Trial after trial: General processing consequences as a function of repetition and change in multidimensional sound. *The Quarterly Journal of Experimental Psychology*, 63, 1770-1788.

- Dyson, B. J. (2011). *Memorial structure in sound: Effects of space, time and preparation on object-based organisation*. Unpublished manuscript.
- Dyson, B. J., & Ishfaq, F. (2008). Auditory memory can be object-based. *Psychonomic Bulletin and Review*, 15, 409-412.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7, 1-14
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10, 1-12.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458-467.
- Fujisaki, W., Shimojo, S., Kashino, S. & Nishida, S. (2004). Recalibration of audio-visual simultaneity. *Nature Neuroscience*, 7, 773-778.
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics*, 68, 1191-1203.
- Gevers, W., & Lammertyn, J. (2005). The hunt for SNARC. *Psychology Science*, 47, 10–21.
- Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York: Wiley.
- Hossain, S. (2011). *Shapes and sounds: an exploration of audiovisual crossmodality*. Unpublished manuscript, University of Texas at Dallas, Dallas, TX.

- Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, 10, 638-644.
- Kawahara, J.-I. (2007). Auditory-visual contextual cuing effect. *Perception & Psychophysics*, 69, 1399-1408.
- Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, 46, 145-149.
- Kim, R., Seitz, A., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE*, 3, e1532.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35-B42.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioural performance. *Experimental Brain Research*, 158, 405-414.
- Leboe, L. C., & Mondor, T. A. (2007). Item-specific congruency effects in nonverbal auditory Stroop. *Psychological Research*, 71, 568-575.
- Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology*, 13, 384-394.
- Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, 50, 125-145.

- Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Test of the semantic coding hypothesis. *Perception*, 28, 903-923.
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *American Psychology Society*, 10, 61-66.
- Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 69-79.
- Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics*, 48, 169-178.
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, and Behavioural Neuroscience*, 4, 133-136.
- Molholm, S., Ritter, W., Javitt, D. C. & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, 14, 452-465.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism, *Cognitive Brain Research*, 17, 154-163.
- Robinson, C. W., & Sloutsky, V. M. (2007). Visual statistical learning: Getting some help from the auditory modality. Paper session presented at the meeting of Cognitive Science Society, Nashville, TN.

- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Sagiv, N., & Ward, J. (2006). Crossmodal interactions: Lessons from synesthesia. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J.-M. Alonso & P. U. Tse (Eds.). *Visual perception: Part 2. Fundamentals of awareness, multi-sensory integration and higher-order period*. (Progress in Brain Research, 155, pp. 263-275. London: Elsevier Science.
- Sandhu, R., & Dyson, B. J. (2012). Re-evaluating visual and auditory dominance through modality switching costs and congruency analyses. *Acta Psychologica*, 140, 111-118.
- Sandhu, R., & Dyson, B. J. (2013). Modality and task switching interactions using bi-modal and bivalent stimuli. *Brain and cognition*, 82, 90-99.
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268, 889-893.
- Schiffman, H. R. (2001). *Sensation and perception*. New York, NY: John Wiley & Sons, Inc.
- Shams, S., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, 408, 788.
- Shams, S., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Science*, 12, 411-417.

- Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, 24, 99-142.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, Psychophysics*, 73, 971-995.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neurophysiologica*, 45, 598-607.
- Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology*, 30, 1131-1142.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 40, 201-237.
- van den Bos, E., Christiansen, M., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, 67, 507-520.
- Walk, A. M., & Conway, C. M. (2011). Multisensory statistical learning: Can associations between perceptual categories be acquired? Paper session presented at the meeting of Cognitive Science Society, Boston, MA.

- Walker, P., & Walker, L. (2012). Size-brightness correspondence: Crosstalk and congruity among dimensions of connotative meaning. *Attention, Perception, Psychophysics*, DOI 10.3758/s13414-012-0297-9.
- Wallach, H., Newman, E. B., & Rosenweig, M. R. (1949). The precedence effect in sound localization. *American Journal of Psychology*, 62, 315-336.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638-667.
- Wilbiks, J., & Dyson, B. J. (in review) Effects of temporal asynchrony and spatial congruency on competitive audio-visual causality.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 1-7.
- Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, 193, 603-614.

## Tables and Figures

**Table 1      Summary of Predictive Values**

Baseline		Intra-modal		Inter-modal	
First	Second	First	Second	First	Second
A1	A2	A1 →	<b>A2</b>	A1 →	A2
	V1		V1		<b>V1</b>
A2	A1	A2 →	<b>A1</b>	A2 →	A1
	V2		V2		<b>V2</b>
V1	V2	V1 →	<b>V2</b>	V1 →	V2
	A1		A1		<b>A1</b>
V2	V1	V2 →	<b>V1</b>	V2 →	V1
	A2		A2		<b>A2</b>

\*Stimulus attribute dimensions in First column have predictive value only for bolded stimulus attribute dimensions in Second column.



**Table 2**      **Median RTs and Error Rates for All Cells in Pilot Study**

Condition	Same-Modality				Different-Modality			
	First		Second		First		Second	
	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)
Baseline	933 (62)	7.29 (1.72)	759 (63)	9.63 (2.20)	892 (54)	7.94 (2.06)	851 (61)	10.42 (2.13)
Intra-modal	856 (61)	5.99 (1.38)	734 (68)	8.98 (2.36)	858 (63)	6.51 (1.91)	814 (69)	9.90 (1.65)
Inter-modal	904 (61)	9.64 (1.67)	774 (62)	9.90 (2.37)	913 (64)	8.20 (2.02)	878 (63)	11.72 (2.19)

**Table 3**      **Summary of ANOVA results for RTs in Pilot Study**

Effect	df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
C	2, 22	2.372	36883	.117	.177
M	1, 11	4.725	60521	.052	.300
<b>R</b>	<b>1, 11</b>	<b>11.864</b>	<b>298067</b>	<b>.005</b>	<b>.519</b>
C x M	1.353, 14.878	1.551	3030	.240	.124
C x R	2, 22	.460	2521	.637	.040
M x R	1, 11	4.739	93044	.052	.301
C x M x R	2, 22	2.545	945	.101	.188

Note: Statistical significance in bold. C = Condition, M = Modality Judgment, R = Response Order.

**Table 4** Summary of ANOVA results for Error Rates in Pilot Study

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
C	2, 22	2.134	22.910	.142	.162
M	1, 11	.723	14.665	.413	.062
<b>R</b>	<b>1, 11</b>	<b>8.175</b>	<b>27.427</b>	<b>.016</b>	<b>.426</b>
C x M	2, 22	.056	19.303	.945	.005
C x R	2, 22	.379	13.605	.689	.033
M x R	1, 11	.923	15.442	.357	.077
C x M x R	2, 22	1.345	6.708	.281	.109

Note: Statistical significance in bold. C = Condition, M = Modality Judgment, R = Response Order.

**Table 5**      **Median RTs and Error Rates for All Cells in Experiment 1**

Condition	Same-Modality				Different-Modality			
	First		Second		First		Second	
	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)
Baseline	921 (33)	9.05 (1.18)	747 (37)	10.01 (1.11)	920 (34)	8.98 (1.09)	894 (29)	15.82 (1.58)
Intra-modal	903 (32)	7.75 (1.09)	734 (31)	7.47 (.90)	899 (33)	8.01 (.88)	880 (37)	15.26 (1.24)
Inter-modal	934 (39)	8.72 (1.25)	760 (37)	13.59 (2.12)	877 (31)	9.18 (1.61)	838 (34)	13.24 (1.52)

**Table 6** Summary of ANOVA results for RTs in Experiment 1

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
C	2, 46	.509	9836	.605	.022
<b>M</b>	<b>1, 23</b>	<b>16.070</b>	<b>191168</b>	<b>&lt;.001</b>	<b>.426</b>
<b>R</b>	<b>1, 23</b>	<b>38.341</b>	<b>18818</b>	<b>&lt;.001</b>	<b>.625</b>
<b>C x M</b>	<b>2, 46</b>	<b>8.831</b>	<b>3424</b>	<b>&lt;.001</b>	<b>.277</b>
C x R	2, 46	.204	5040	.816	.009
<b>M x R</b>	<b>1, 23</b>	<b>28.728</b>	<b>12955</b>	<b>&lt;.001</b>	<b>.555</b>
C x M x R	2, 46	.103	3798	.903	.004

Note: Statistical significance in bold. C = Condition, M = Modality Judgment, R = Response Order.

**Table 7** Summary of ANOVA results for Error Rates in Experiment 1

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
C	2, 46	1.887	36	.163	.076
<b>M</b>	<b>1, 23</b>	<b>22.565</b>	<b>17</b>	<b>&lt;.001</b>	<b>.495</b>
<b>R</b>	<b>1, 23</b>	<b>45.071</b>	<b>25</b>	<b>&lt;.001</b>	<b>.662</b>
<b>C x M</b>	<b>2, 46</b>	<b>5.827</b>	<b>17</b>	<b>.006</b>	<b>.202</b>
C x R	2, 46	.170	34	.844	.007
<b>M x R</b>	<b>1, 23</b>	<b>17.607</b>	<b>18</b>	<b>&lt;.001</b>	<b>.434</b>
<b>C x M x R</b>	<b>2, 46</b>	<b>6.611</b>	<b>18</b>	<b>.003</b>	<b>.223</b>

Note: Statistical significance in bold. C = Condition, M = Modality Judgment, R = Response Order.

**Table 8      Transfer of Mappings**

Experiment 1		Experiment 2	
Shape	Square	Shape	Circle
	Diamond		Cloud
Depth	2D	Colour	White
	3D		Grey
Pitch	Low	Pitch	Low
	High		High
Warble	No	Warble	No
	Yes		Yes

**Table 9**      **Median RTs and Error Rates for All Cells in Experiment 2**

Condition	Same-Modality				Different-Modality			
	First		Second		First		Second	
	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)
Baseline	988 (40)	7.16 (1.27)	808 (45)	9.65 (1.44)	992 (41)	7.75 (1.42)	972 (30)	12.72 (1.68)
Intra-modal	972 (47)	6.58 (1.09)	807 (50)	6.32 (1.03)	981 (53)	6.97 (1.19)	971 (33)	11.70 (1.48)
Inter-modal	960 (46)	7.75 (1.41)	782 (37)	11.09 (1.36)	983 (42)	7.75 (.91)	951 (39)	9.90 (1.15)



**Table 10** Summary of ANOVA results for RTs in Experiment 2

Effect	df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
C	2, 46	.384	28571	.683	.016
<b>M</b>	<b>1, 23</b>	<b>45.828</b>	<b>12393</b>	<b>&lt;.001</b>	<b>.666</b>
<b>R</b>	<b>1, 23</b>	<b>20.279</b>	<b>33835</b>	<b>&lt;.001</b>	<b>.469</b>
C x M	2, 46	.256	4147	.776	.011
C x R	1.457, 33.518	.280	9283	.687	.012
<b>M x R</b>	<b>1, 23</b>	<b>32.583</b>	<b>13124</b>	<b>&lt;.001</b>	<b>.586</b>
C x M x R	2, 46	.074	4423	.928	.003

Note: Statistical significance in bold. C = Condition, M = Modality Judgment, R = Response Order.

**Table 11** Summary of ANOVA results for Error Rates in Experiment 2

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
C	2, 46	1.237	46	.300	.051
<b>M</b>	<b>1, 23</b>	<b>7.346</b>	<b>18</b>	<b>.012</b>	<b>.242</b>
<b>R</b>	<b>1, 23</b>	<b>29.845</b>	<b>20</b>	<b>&lt;.001</b>	<b>.565</b>
<b>C x M</b>	<b>2, 46</b>	<b>6.341</b>	<b>12</b>	<b>.004</b>	<b>.216</b>
C x R	2, 46	.950	14	.394	.040
<b>M x R</b>	<b>1, 23</b>	<b>5.394</b>	<b>15</b>	<b>.029</b>	<b>.190</b>
<b>C x M x R</b>	<b>2, 46</b>	<b>4.297</b>	<b>14</b>	<b>.019</b>	<b>.157</b>

Note: Statistical significance in bold. C = Condition, M = Modality Judgment, R = Response Order.

**Table 12** Summary of ANOVA results for RTs for Cross-Experiment Analysis

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta\rho^2$
E	1, 46	2.432	303780	.126	.05
C	2, 92	.790	23954	.457	.017
C x E	2, 92	.078	23954	.925	.002
<b>M</b>	<b>1, 46</b>	<b>60.110</b>	11796	<b>&lt;.001</b>	<b>.567</b>
<b>M x E</b>	<b>1, 46</b>	<b>4.243</b>	<b>11796</b>	<b>.045</b>	<b>.084</b>
<b>R</b>	<b>1, 46</b>	<b>53.461</b>	<b>26327</b>	<b>&lt;.001</b>	<b>.538</b>
R x E	1, 46	.008	26327	.927	.000

Statistical learning and congruency in multi-modal binding

C x M	2, 92	2.653	3786	.076	.055
<b>C x M x E</b>	<b>2, 92</b>	<b>5.614</b>	3786	<b>.005</b>	<b>.109</b>
C x R	1.758, 80.876	.474	5903	.624	.010
C x R x E	1.758, 80.876	.021	5903	.980	.000
<b>M x R</b>	<b>1, 46</b>	<b>61.262</b>	962	<b>&lt;.001</b>	<b>.571</b>
M x R x E	1, 46	.074	962	.787	.002
C x M x R	2, 92	.158	69	.854	.003
C x M x R x E	2, 92	.017	69	.983	.000

---

Note: Statistical significance in bold. E = Experiment, C = Condition, M = Modality Judgment, R = Response Order.

**Table 13** Summary of ANOVA results for Error Rates for Cross-Experiment Analysis

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta\rho^2$
E	1, 46	1.856	255	.180	.039
C	2, 92	2.990	46	.055	.061
C x E	2, 92	.055	46	.946	.001
<b>M</b>	<b>1, 46</b>	<b>27.548</b>	<b>18</b>	<b>&lt;.001</b>	<b>.375</b>
M x E	1, 46	1.816	18	.184	.038
<b>R</b>	<b>1, 46</b>	<b>74.714</b>	<b>20</b>	<b>&lt;.001</b>	<b>.619</b>
R x E	1, 46	1.735	20	.194	.036
<b>C x M</b>	<b>2, 92</b>	<b>12.022</b>	<b>12</b>	<b>&lt;.001</b>	<b>.207</b>

Statistical learning and congruency in multi-modal binding

C x M x E	2, 92	.055	12	.946	.001
C x R	2, 92	.497	14	.610	.011
C x R x E	2, 92	.312	14	.733	.007
<b>M x R</b>	<b>1, 46</b>	<b>21.839</b>	<b>15</b>	<b>&lt;.001</b>	<b>.322</b>
M x R x E	1, 46	2.458	15	.124	.051
<b>C x M x R</b>	<b>2, 92</b>	<b>10.756</b>	<b>14</b>	<b>&lt;.001</b>	<b>.190</b>
C x M x R x E	2, 92	.461	14	.632	.010

---

Note: Statistical significance in bold. E = Experiment, C = Condition, M = Modality Judgment, R = Response Order.

**Table 14**      **Median RTs and Error Rates by Congruency in Baseline Condition in Experiment 2**

Number of Congruent Pairings	Same-Modality		Different-Modality	
	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)
None	868 (39)	8.33 (1.43)	969 (42)	9.77 (1.76)
One	874 (33)	7.62 (1.11)	934 (29)	8.92 (1.09)
Two	853 (32)	8.46 (1.68)	946 (25)	10.29 (1.75)

**Table 15** Summary of ANOVA Results for RTs for Congruencies in Baseline Condition in Experiment 2

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
<b>M</b>	<b>1, 23</b>	<b>11.437</b>	<b>21337</b>	<b>.003</b>	<b>.332</b>
C	2, 46	1.484	10004	.237	.061
<b>M x C</b>	<b>2, 46</b>	<b>4.095</b>	<b>5390</b>	<b>.023</b>	<b>.151</b>

Note: Statistical significance in bold. M = Modality, C = Congruency.



**Table 16** Summary of ANOVA Results for Error Rates for Congruencies in Baseline Condition in Experiment 2

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta p^2$
M	1, 23	2.668	26	.116	.104
C	<b>2, 46</b>	<b>8.830</b>	<b>32</b>	<b>&lt;.001</b>	<b>.277</b>
M x C	2, 46	.049	26	.952	.002

Note: Statistical significance in bold. M = Modality, C = Congruency.

**Table 17**      **Median RTs and Error Rates by Congruency in Baseline Condition in Experiment 2 on Trials with Congruent-Association Prompts**

Number of Congruent Pairings	Same-Modality		Different-Modality	
	RT (SE)	Errors (SE)	RT (SE)	Errors (SE)
None	907 (62)	10.16 (2.00)	1044 (79)	8.59 (2.15)
One	906 (47)	7.42 (1.41)	985 (38)	7.94 (1.52)
Two	887 (47)	7.03 (2.18)	989 (40)	9.11 (2.29)

**Table 18**      **Summary of ANOVA Results for RTs for Congruencies in Baseline Condition in Experiment 2 on Trials with Congruent-Association Prompts**

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta\rho^2$
M	1, 11	4.723	35090	.052	.300
C	2, 22	1.437	13781	.259	.116
M x C	2, 22	2.340	6902	.120	.175

Note: M = Modality, C = Congruency.

**Table 19** Summary of ANOVA Results for Error Rates for Congruencies in Baseline Condition in Experiment 2 on Trials with Congruent-Association Prompts

Effect	Df	<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta\rho^2$
M	1, 11	2.037	15	.181	.156
<b>C</b>	<b>2, 22</b>	<b>3.548</b>	<b>40</b>	<b>.046</b>	<b>.244</b>
M x C	2, 22	.039	31	.961	.004

Note: Statistical significance in bold. M = Modality, C = Congruency.

**Figure 1**     **Stimuli**

Experiment 1



HIGH

FM



HIGH

NO FM



LOW

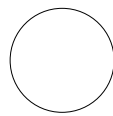
FM



LOW

NO FM

Experiment 2



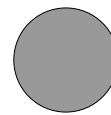
HIGH

NO FM



HIGH

FM



LOW

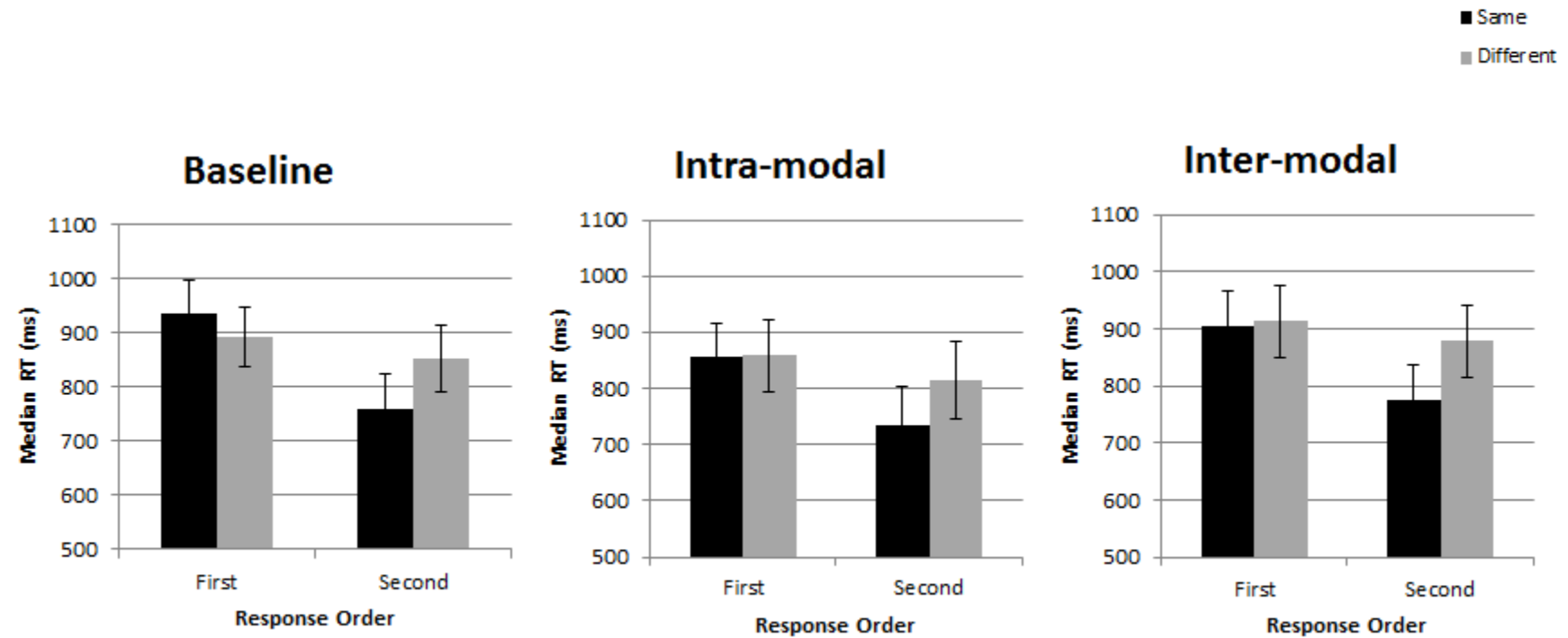
NO FM



LOW

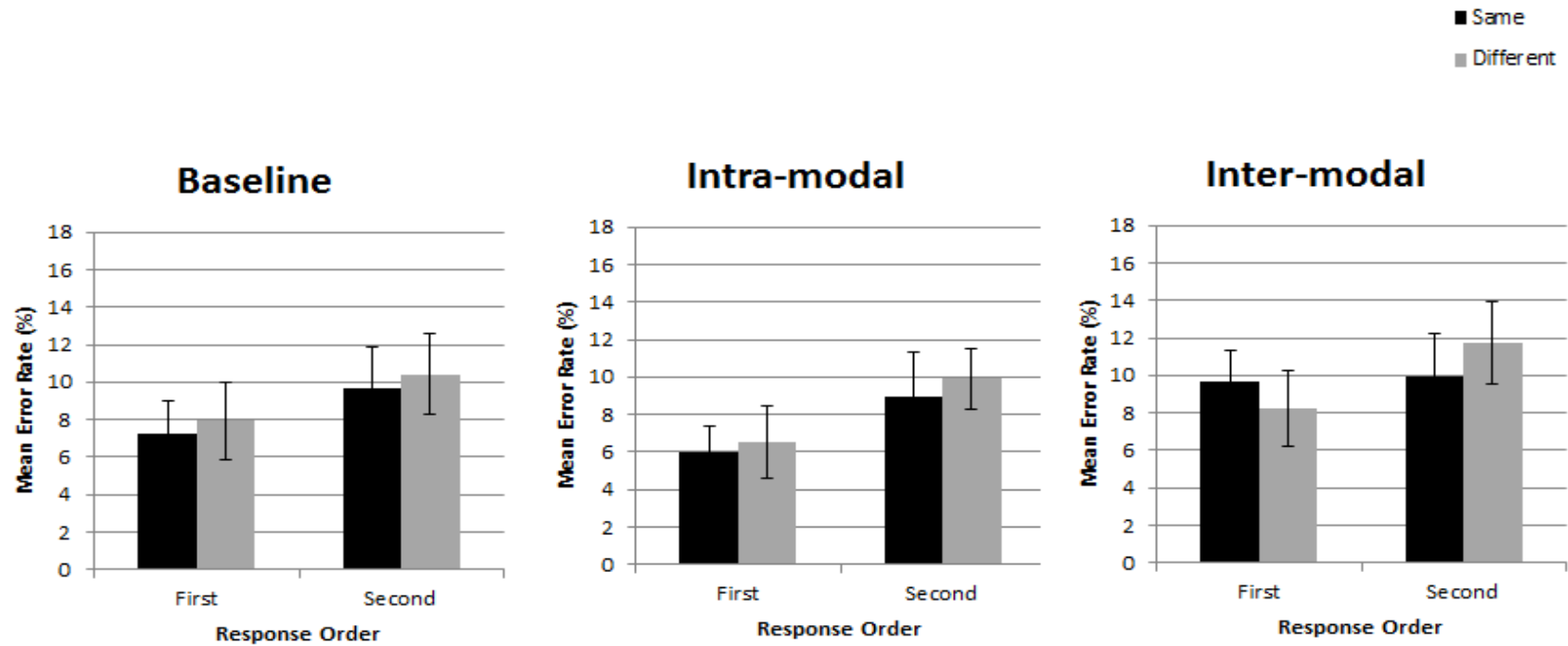
FM

**Figure 2**      **Median Reaction Time for All Conditions in Pilot Study**



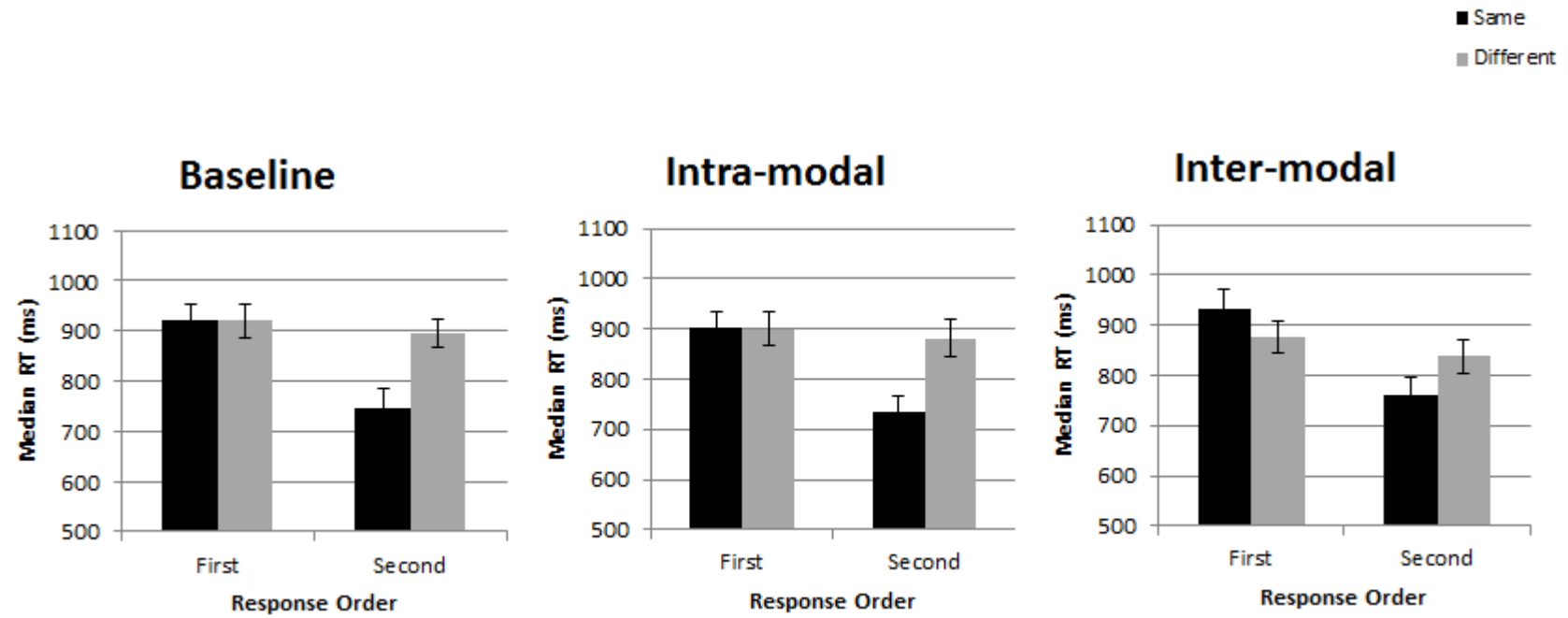
Error Bars: +/- 1 SE

**Figure 3**      **Mean Error Rates for All Conditions for Pilot Study**



Error Bars: +/- 1 SE

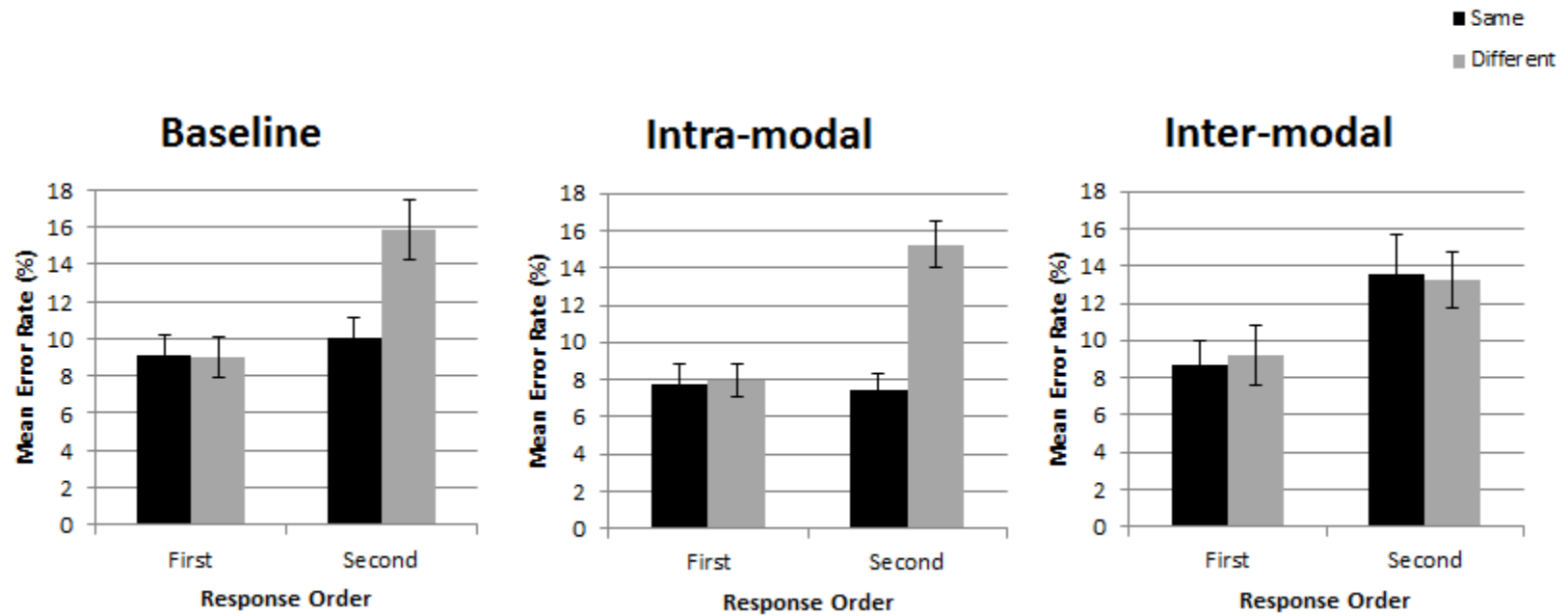
**Figure 4**      **Median Reaction Time for All Conditions in Experiment 1**



Error Bars: +/- 1 SE

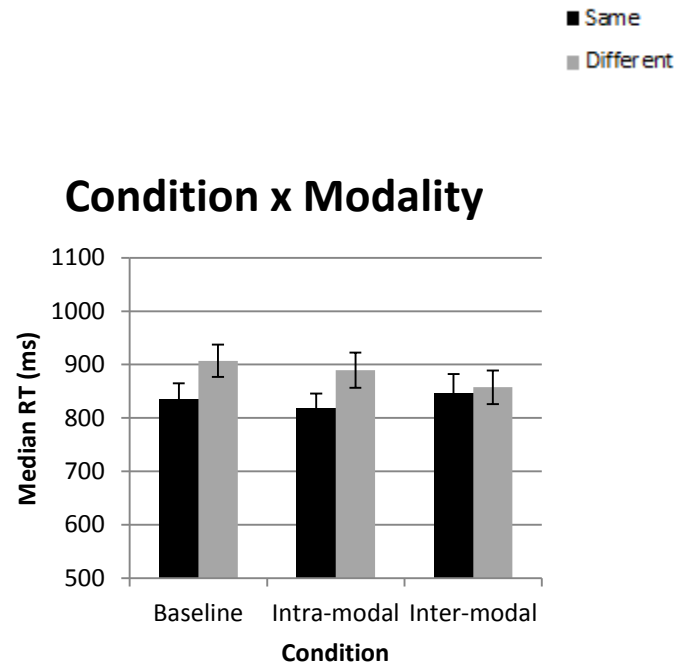


**Figure 5** Mean Error Rate for All Conditions in Experiment 1



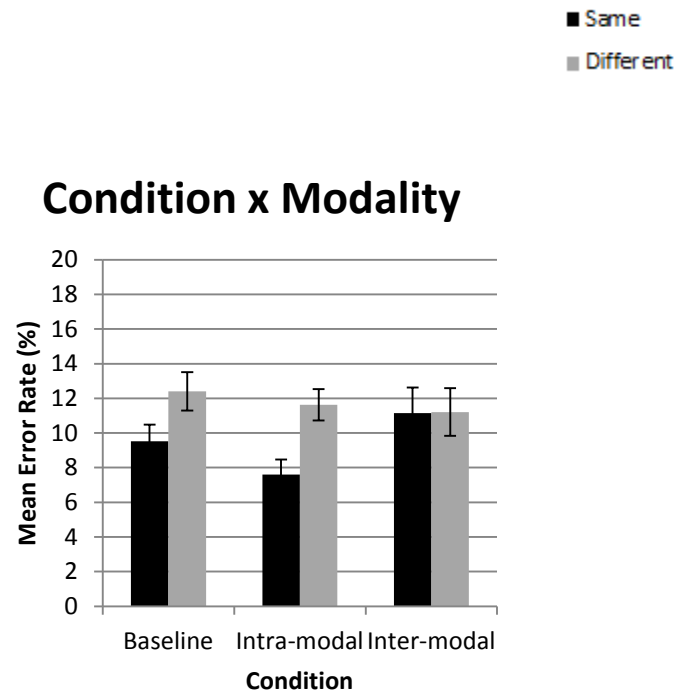
Error Bars: +/- 1 SE

**Figure 6      Median Reaction Time for Condition x Modality in Experiment 1**



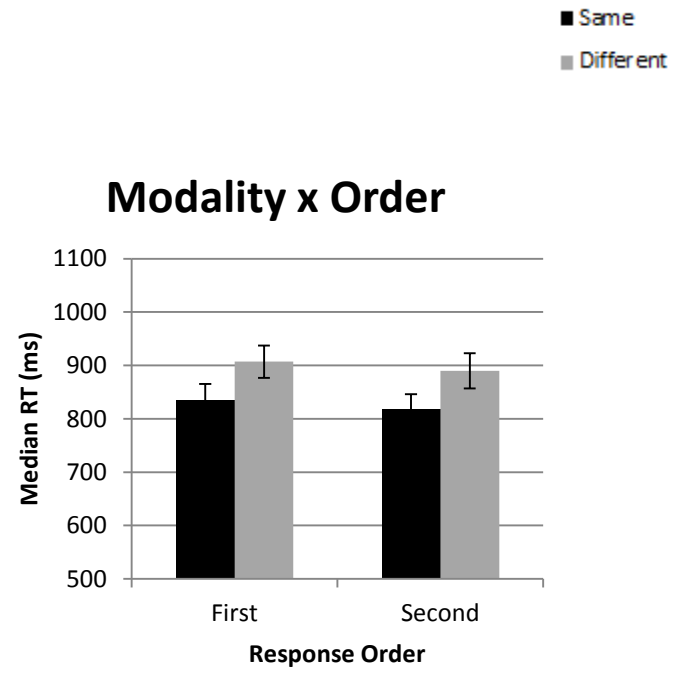
Error Bars: +/- 1 SE

**Figure 7 Mean Error Rate for Condition x Modality in Experiment 1**



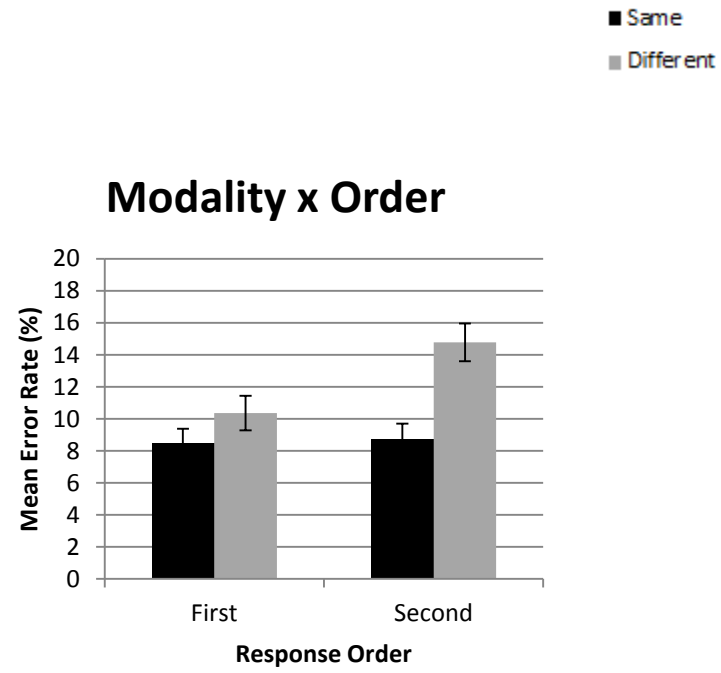
Error Bars: +/- 1 SE

**Figure 8      Median Reaction Time for Modality x Order in Experiment 1**



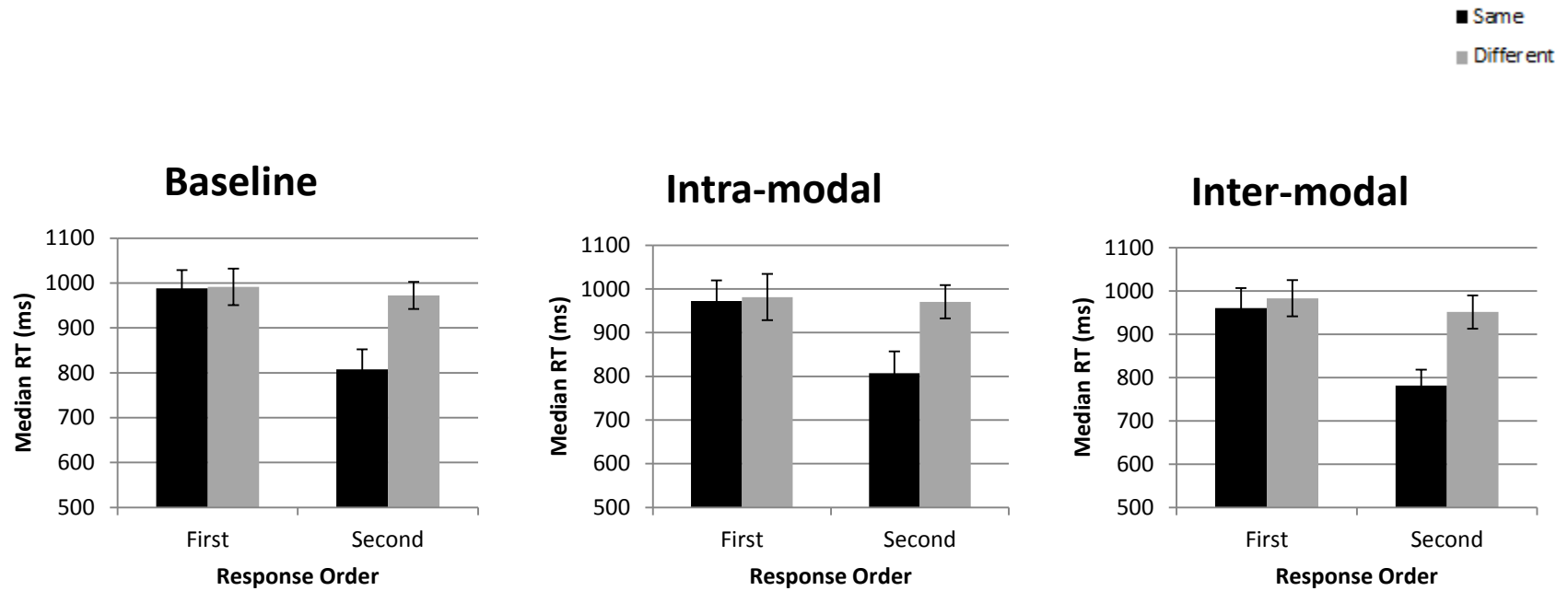
Error Bars:  $\pm 1$  SE

**Figure 9**      **Mean Error Rate for Modality x Order in Experiment 1**



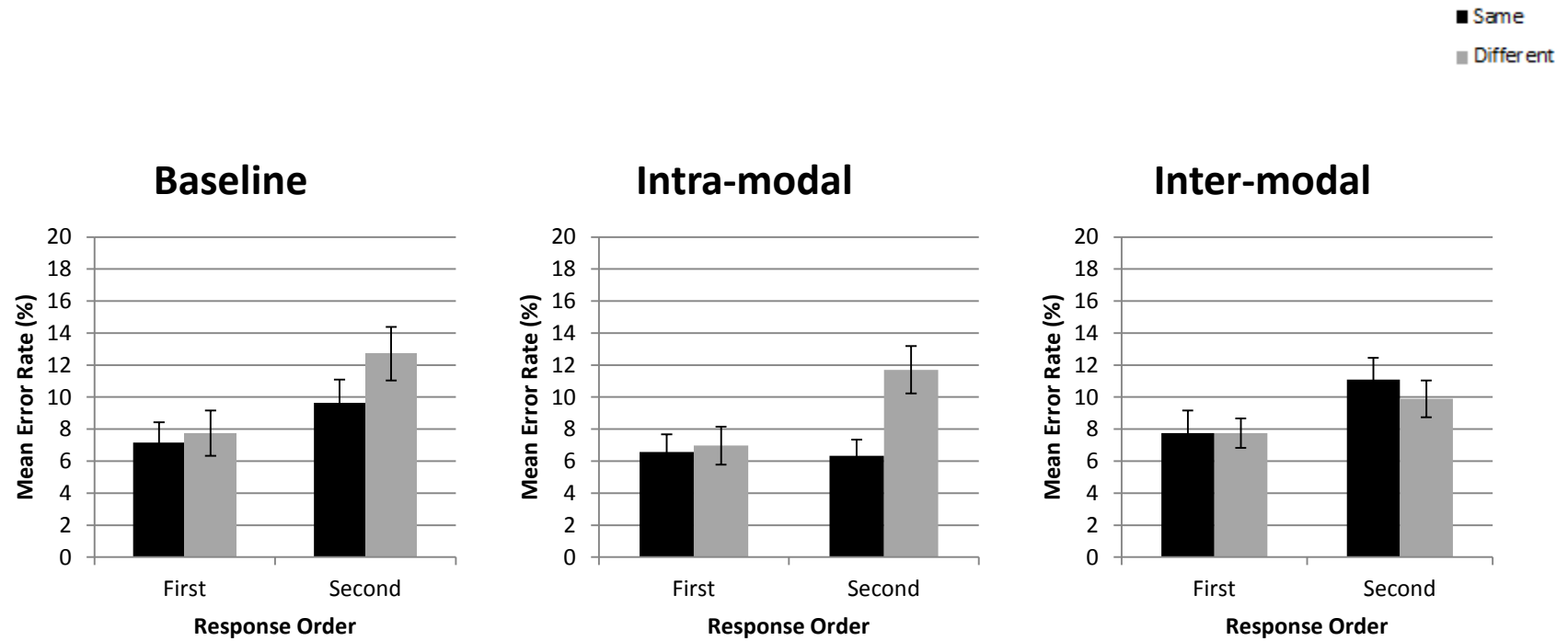
Error Bars:  $\pm 1$  SE

**Figure 10 Median Reaction Time for All Conditions in Experiment 2**



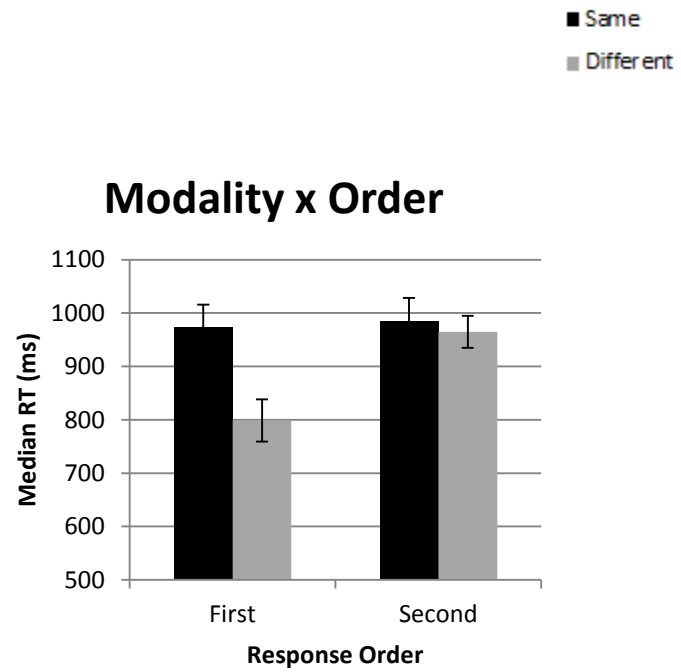
Error Bars: +/- 1 SE

**Figure 11** Mean Error Rate for All Conditions in Experiment 2



Error Bars: +/- 1 SE

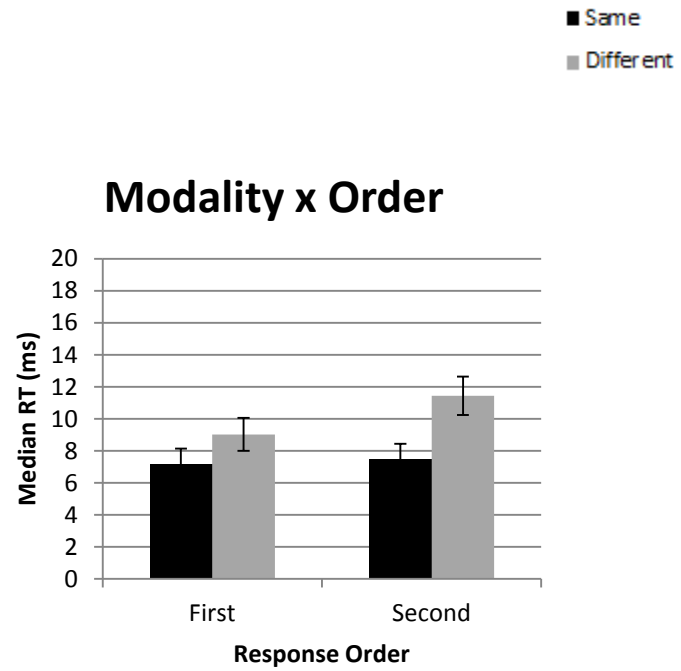
**Figure 12     Median Reaction Time for Modality and Order in Experiment 2**



Error Bars:  $\pm 1$  SE

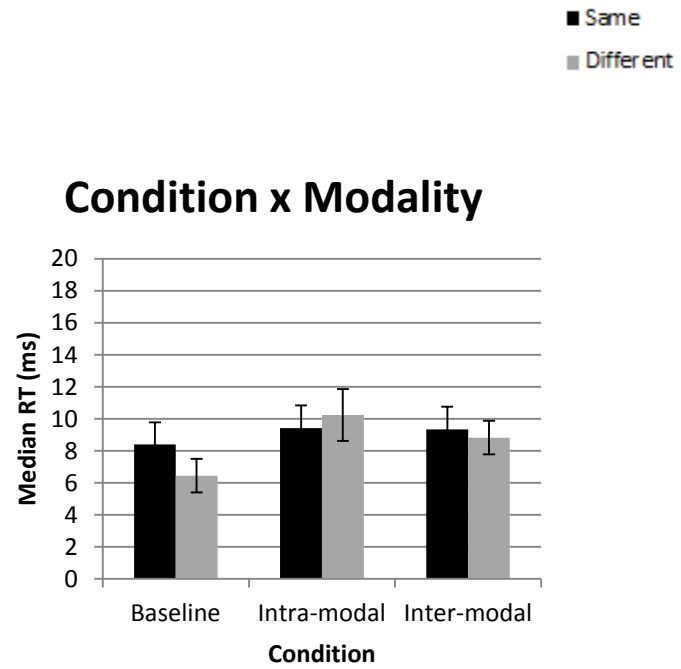


**Figure 13**      **Mean Error Rate for Modality x Order in Experiment 2**



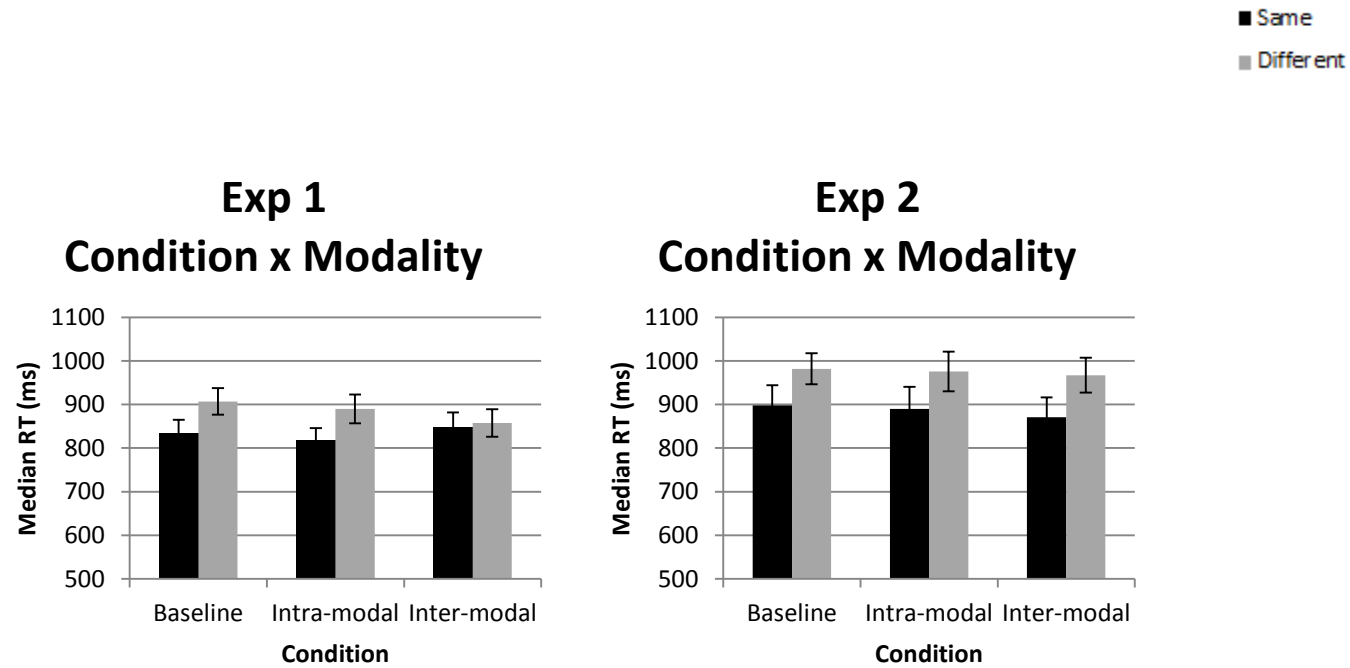
Error Bars: +/- 1 SE

**Figure 14** Mean Error Rate for Condition x Modality in Experiment 2



Error Bars: +/- 1 SE

**Figure 15** Median Reaction Time for Condition x Modality x Experiment in Cross-Experiment Analysis



Error Bars: +/- 1 SE

## Ryerson University

## Consent Agreement

**STIMULUS AND PROBABILITY CONSTRAINTS IN THE  
DEVELOPMENT OF MULTI-MODAL OBJECTS**

You are being asked to participate in a research study. Before you give your consent to be a volunteer, it is important that you read the following information and ask as many questions as necessary to be sure you understand what you will be asked to do.

**Investigators:** Ben Dyson (faculty), Zara Chan (MA student)

**Purpose of the Study:** This one hour lab-based cognitive study aims to examine how the integration of auditory and visual information can be promoted. We are hoping to test 24 healthy adults across different age groups, and wish to use only those individuals who self-report as having normal hearing and vision, or corrected hearing and vision with the use of hearing aids, glasses/contact lenses or other devices.

**Description of the Study:** The study will take place in the HEAR Lab, located in the Psychology Research and Training Centre at 105 Bond Street, unless otherwise stated. The experiment will take 1 hour to complete. Prior to the study, you will have the study explained to you and the opportunity to take part in a practice block so you are familiar with the procedure. You will be given the chance to ask any questions you may have regarding the study, prior to reviewing the consent agreement. During the study, you are invited to provide demographic information (age, handedness, gender). After the study, you will be fully debriefed as to the purpose of the study, and given a further opportunity to ask questions.

At each trial, you will be presented with both a visual object and an auditory object. The visual object might be composed of a certain shape (e.g., circle or square) in a certain colour (e.g., red or green), whereas the auditory object might be composed of a certain pitch (e.g., high or low) which may or may not have a ‘warbling’ nature (e.g., yes or no). **The experimenter will clarify how the visual and auditory objects vary in this particular experiment.** After stimulus presentation, you will be asked two questions about what you just heard and saw. Sometimes the questions will ask about two properties from the same object (for example, what was the colour of the visual object and what was the shape of the visual object?). Sometimes the questions will ask about two properties from different objects (for example, what was the colour of the visual object and what was the pitch of the auditory object?). Fast and accurate responses regarding what you saw and heard are preferred.

**What is Experimental in this Study:** Previous research has been interested in how audio and visual information is put together, and this study aims to assess how the combination of certain stimulus properties might enhance that integration.

**Risks or Discomforts:** There are no known long-term risks associated with behavioural testing of the manner proposed. One short-term risk is fatigue. Effects of fatigue will be offset by providing you with the opportunity to take breaks in-between blocks of trials. *You may discontinue participation, either temporarily or permanently at any time for any reason.*

**Benefits of the Study:** This research will contribute to an understanding of how relationships between sound and vision are established and how people put information from different sense together. No individual direct benefit can be guaranteed by the researchers.

**Confidentiality:** Confidentiality will be maintained in all aspects of data dissemination. Only identifying information (name) will appear on this consent form. For all other aspects of the study, a unique numeric ID will be assigned for each participant. Names and IDs will not be matched. Original paper records will be stored in a locked file cabinet and electronic records will be stored on password-protected computers. All data will be stored for a minimum of 1 year after collection. Data is typically retained for 5 years after publication of the study with hardcopy data will be destroyed by confidential shredding; electronic data will be destroyed by deletion. Participants have the option of reviewing and / or removing all of their data from the study, if the request is made immediately after the study.

**Compensation For Participation:** Three different compensation schemes are offered. Payment can be offered for experimental participation at the rate of \$10 per hour or for one course credit, even if either the participant or experimenter chooses to discontinue the study. For certain individuals, course credit is available as compensation, awarded either on the basis of participation or a walk-through in which the participant can take part in the study but not submit their data. Please indicate which compensation you require:

☐ MONEY

☐ COURSE CREDIT  
(PARTICIPATION)

☐ COURSE CREDIT  
(WALKTHROUGH)

**Voluntary Nature of Participation:** Participation in this study is voluntary. Your choice of whether or not to participate will not influence your future relations with Ryerson University. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time without penalty or loss of benefits to which you are allowed. At any particular point in the study, you may refuse to answer any particular question or stop participation altogether.

**Questions about the Study:** If you have any questions about the research now, please ask. If you have questions later about the research, you may contact:

Ben Dyson, Ph.D.  
Principal Investigator  
ben.dyson@psych.ryerson.ca  
416-979-5000 x2063

If you have questions regarding your rights as a human subject and participant in this study, you may contact the Ryerson University Research Ethics Board for information.

Research Ethics Board  
c/o Office of the Vice President, Research and Innovation  
Ryerson University, 350 Victoria Street  
Toronto, ON, M5B 2K3, Canada  
001 416-979-5042

**Agreement:** Your signature below indicates that you have read the information in this agreement and have had a chance to ask any questions you have about the study. Your signature also indicates that you agree to be in the study and have been told that you can change your mind and withdraw your consent to participate at any time. You have been given a copy of this agreement. You have been told that by signing this consent agreement you are not giving up any of your legal rights.

*Informed consent for study participation*

\_\_\_\_\_  
Name of Participant (please print)

\_\_\_\_\_  
Signature of Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature of Investigator

\_\_\_\_\_  
Date

## **Stimulus and probability constraints in the development of multi-modal objects**

Dear Participant:

Thank you very much for your participation in our study. Your time and commitment to psychological research at Ryerson University is very much appreciated.

The study you took part in will contribute to ongoing auditory and visual research conducted in the H.E.A.R Lab. Our lab is dedicated to designing and implementing research studies that will help us better understand how the brain represents what we hear and see, and how this information is integrated.

The particular study you took part in was designed to assess how individuals learn to integrate information from the different senses. We are assuming that it is initially easy to remember two pieces of information from the same sense (e.g., vision or audition). However, we eventually learn that certain sounds go with certain sights because of two factors: 1) the sounds and sights co-occur (e.g., you rectangular mobile phone has a certain ring) and 2) the sounds and sights somehow ‘fit together’ better (e.g., low sounds appear of the left side of visual space, like on a piano). By examining data across these different experiments, we will look to see how much better people become at remembering two pieces of information from different senses (e.g., vision and audition) as a result of the 1) co-occurrence and 2) stimulus fit. In this way, we will begin to understand how the eyes and the ears communicate in the brain.

If you have any questions regarding your participation in this study, or would like to receive information about the results once they are available, feel free to contact Dr. Ben Dyson. We would be happy to provide you with the overall findings of our study.

Finally, if you are interested in taking part and learning more about visual and auditory perception research in the H.E.A.R Lab, feel free to contact Dr. Dyson.

---

Zara Chan  
MA Student  
Ryerson University  
z5chan@psych.ryerson.ca

Dr. Ben Dyson  
Assistant Professor  
Ryerson University  
[ben.dyson@psych.ryerson.ca](mailto:ben.dyson@psych.ryerson.ca)