

1-1-2011

Multimodal Integration and Expertise

Rajwant Sandhu
Ryerson University

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Cognition and Perception Commons](#)

Recommended Citation

Sandhu, Rajwant, "Multimodal Integration and Expertise" (2011). *Theses and dissertations*. Paper 1715.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact bcameron@ryerson.ca.

MULTIMODAL INTEGRATION AND EXPERTISE

by

Rajwant Sandhu,
Hon B.Sc., University of Waterloo, 2004

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, 2011

© Rajwant Sandhu 2011

Ryerson University

Author's Declaration

I hereby declare that I am the sole author of this thesis or dissertation.

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

Signature

I further authorize Ryerson University to reproduce this thesis or dissertation 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.

Signature

MULTIMODAL INTEGRATION AND EXPERTISE

By Rajwant Sandhu

Ryerson University, Master of Arts, Psychology, 2011

Abstract

Multi-modal integration often results in one modality dominating sensory perception. Such dominance is influenced by task demands, processing efficiency, and training. I assessed modality dominance between auditory and visual processing in a paradigm controlling for the first two factors while manipulating the third. In a uni-modal task auditory and visual processing was equated per individual participant. Pre and post training, participants completed a bimodal selective attention task where the relationship between relevant and irrelevant information, and the task-relevant modality changed across trials. Training in one modality was provided between pre and post-training tasks. Training resulted in non-specific speeding post-training. Pre-training, visual information impacted auditory responding more than vice versa and this pattern reversed following training, implying visual dominance pre, and auditory dominance post-training. Results suggest modality dominance is flexible and influenced by experimental design and participant abilities. Research should continue to uncover factors leading to sensory dominance by one modality.

Acknowledgements

I would like to sincerely thank my advisor Ben Dyson for all of the time and effort he put towards assisting me with this project. I would also like to thank my thesis supervisory committee, consisting of Frank Russo, Lixia Yang as well as my external examiner, Todd Girard, for helpful comments and suggestions. Finally, I am grateful to my husband for his tireless support during this process.

Table of Contents

Author's Declaration	ii
Abstract	iii
Acknowledgements.....	iv
List of Tables	vii
List of Figures	viii
Introduction	1
Feature binding	2
Intramodal feature binding.....	3
Intermodal feature binding.....	7
Factors affecting MMI.....	9
Speed of processing	10
Task Requirements.....	12
Attentional load	14
Expertise.....	15
Objectives of the current proposal	17
Methods.....	21
Participants	21
Design.....	23
Procedures	26
Results.....	27
Data Exclusion	27
Uni-modal Baseline	27

RT's.....	28
Error Rates	29
Unblocked Filter.....	29
RT 's.....	31
Error Rates	32
Discussion.....	37
Audio-visual binding.....	38
Modality Dominance.....	42
Uni-modal Responding.....	46
Future Research	47
General Conclusions.....	49
References	51

List of Tables

Table 1. Uni-baseline task reaction times and error rates ANOVA table.....	28
Table 2. Filter task reaction times and error rates ANOVA table.	30
Table 3. Table of Follow-up ANOVAs for training group and irrelevance by filter block in reaction times.....	31
Table 4. Table of follow-up three-way ANOVAs for modality, relevance and irrelevance by filter block in error rate data.	34
Table 5. Table of follow-up ANOVAs on relevance and irrelevance by modality in post-training filter block in error rates.....	37

List of Figures

Figure 1. Schematic representation of experimental design.....	24
Figure 2. Reaction times for uni-modal blocks by training group pre and post-training.	28
Figure 3. Interaction between training group and irrelevance by Filter block (pre and post training).....	32
Figure 4. Interaction between relevance and irrelevance, by modality and filter block.....	33
Figure 5. Interaction between modality and relevance in pre-training filter block.....	35
Figure 6. Relevant repeat by relevant change interaction in pre-training filter block.	35

Introduction

Our senses are constantly bombarded with simultaneous multi-sensory information from multiple sources. Thus, our perceptual system is charged not only with maintaining separation between information obtained from different sensory sources, but also ensuring that only those features belonging to a single source are bound together. While feature binding within a modality has been extensively studied (e.g., Allport, Tipper, & Chmiel 1985; Dyson and Quinlan, 2004; Hall, Pastore, Acker, & Huang 2000; Takegata, Brattico, Tervaniemi, Varyagina, Näätänen, & Winkler 2005; Treisman & Gelade, 1980; Woods, Alain, Diaz, Rhodes & Ogawa, 2001) there exists a paucity of research on feature binding between sensory modalities. This is perhaps due to the fact the factors influencing feature binding between modalities are greater in number than the factors influencing feature binding within modalities. In addition to factors such as attention and relative signal strength (Meredith & Stein, 1987) which influence binding both within and between senses, the relative strengths of the modalities which comprise the multimodal object (i.e., modality appropriateness, Welch and Warren, 1980; modality dominance, Posner, 1976) as well as spatial and temporal coincidence must be considered when examining multimodal integration processes. The current review of literature will first present a discussion of theories focused on how features are bound, both within and between modalities. This will be followed by findings from studies attempting to outline the processes involved in multimodal integration and the various factors that lead to different patterns of performance, namely the chosen stimuli, the task, and the perceptual expertise of participants.

Feature binding

Given what we know about the cortical structure of our sensory systems, namely the fact that different cortical areas are involved in processing different aspects of a stimulus (even within a single sensory system), there must exist a mechanism by which information from different cortical areas is considered with respect to the formation of a unified percept. The question of *how* stimulus features are integrated has been labeled as the binding problem (Treisman, 1996). In terms of vision, the conceptualization of the binding problem can be broken down into three parts; parsing, encoding and structural recognition (Treisman, 1999). Parsing refers to the process by which the relevant to-be bound features are isolated and selected from those belonging to other objects, events or ideas; encoding refers to how the bound features are initially represented and transferred to other areas of the brain for processing; structural recognition refers to the ability to define the relationship between bound features belonging to a single object. The extent to which these principles of visual feature binding apply to other modalities such as audition have been explored, and broad equivalencies regarding segregation and integration (Sussman, 2005) and additional processing related to feature binding (Dyson & Quinlan, 2003) have been observed. Some discrepancies between visual and auditory feature binding have also been noted, such as whether stages of auditory feature activation and integration run in parallel (e.g., Woods & Alain, 1993), in contrast to a more serial account of feature activation and integration in vision (Treisman & Gelade, 1980). Furthermore, whereas in vision, space is thought to be the property that attention must be focused on when binding visual features (e.g., Wolfe, 1994), it has been suggested that frequency, rather than location, must be attended for auditory feature binding (see Woods et al., 2001 for a discussion of Frequency Based Feature Integration Theory [FB-FIT]).

Traditionally feature binding has been studied within a single modality (e.g., Dyson & Quinlan, 2004; Hall et al., 2000; Hommel, 2004; Treisman & Gelade, 1980; Woods & Alain, 2001; Woods et al., 2001; Zmigrod & Hommel, 2009). As such, many of the theories on feature binding are based on features generated from within a modality, traditionally with a greater focus on binding visual features. However, in natural environments most of the objects we encounter are experienced via multiple senses, leading to a more recent interest in feature binding between modalities. The following review of literature will thus first examine theories regarding within-modality feature binding (i.e., intramodal feature binding) and then discuss theories regarding between-modality feature binding (i.e., intermodal feature binding).

Intramodal feature binding

When considering feature binding within the visual modality, Feature Integration Theory (FIT; Treisman & Gelade, 1980) is among the most influential (see Quinlan, 2004, for a review). This theory proposes that primary visual features are processed in parallel, and that each feature is represented by a different feature map. The various feature maps are spatially co-ordinated within a master map, thus allowing an object file to be created, when attention is directed to a specific point in space. The various feature maps are integrated and merged into the master map based on spatio-temporal factors in that features with close spatial and temporal proximity are likely to be integrated into a single object file (Kahneman, Treisman, & Gibbs, 1992; Treisman, 1990; Treisman & Gelade, 1980). Object file theory has been assessed by looking at repetition effects, as specific predictions regarding stimulus repetition and change are set out by this theory (see Treisman, 1980). Object file theory predicts that processing is more efficient on the current trial (trial n) when all stimulus features from the previous trial (trial $n-1$) are repeated (complete repetition), compared to when only some features from trial $n-1$ are repeated (partial change), or

no features from trial $n-1$ are repeated (complete alternation). These predictions are based on the notion that in the case of partial repetition, the object file created at trial $n-1$ will be accessed when some of the features of this file are repeated at trial n , however this object file will no longer be accurate and will thus require some form of updating before an accurate response can be made. In the case of complete alternation, the perceptual system does not access the object file created at trial $n-1$ and thus has to create an entirely new object file, which may also lead to slower response times. However, this theory is unable to account for two consistent patterns of responses: the first being that complete alternation often yields comparable reaction times to complete repetition, and the second being that complete stimulus repetition causes speeding of reaction time (RTs) when the response also repeats, but a slowing of reaction times when the response changes (Hommel, 1998). Thus, the notion of an *object file* has been extended to an *event file* which includes response-related information in addition to stimulus-related information across trials (Hommel, 1998). Thus, event file theory is able to provide a theoretical account for RT patterns on stimulus-response couplings rather than stimulus effects per se. Specifically, event file theory suggests that stimulus and response features are integrated into an event file at trial $n-1$, and that repetition of any of the separable features of the event file created at trial n lead to the retrieval of the recently created event file. In the case of complete stimulus and response repetition the event file is retrieved and accurate, in the case of partial stimulus and response repetition the event file is retrieved but requires updating to reflect the appropriate change(s), and finally in the case of complete stimulus and response change the event file will not be retrieved, but a new file will need to be created (see also Rabbitt, 1992, for a discussion of the similar Bypass Rule).

In line with object and event file theories, visual research consistently demonstrates a same object advantage for memory; that is multiple visual features originating from the same spatial location are more likely to be remembered if they are perceived as belonging to a single object. Conversely, individuals are less accurate when retrieving feature that do not belong to the same object (Duncan, 1984). Recent research suggests that certain auditory stimuli are also registered in terms of objects, by which representations of integrated featural information (templates) are formed (Mondor, Zatorre & Terrio, 1988; see Shinn-Cunningham, 2008, for a review). Hecht, Abbs and Vecera (2008), and, Dyson and Ishfaq (2008) demonstrated a same-object effect in the auditory domain. Specifically Hecht, Abbs and Vecera (2008) found that participants were more accurate at recalling features belonging to the same object, while Dyson and Ishfaq (2008) found that participants were faster at recalling features when they belonged to the same object. Both studies display greater processing efficiency of features belonging to the same compared to different objects. Perhaps due to the growing interest in feature binding in audition (e.g., Hall et al., 2000; Takegata et al., 2005), Zmigrod and Hommel (2009) examined the extent to which visual event file theory could be applied to auditory stimuli. They found similar patterns of results as those obtained from research with visual stimuli. Namely, similar to visual repetition effects, slowing of reaction times during partial repetition relative to complete repetition or alternation were observed with auditory stimuli. The slowing associated with partial repetition provides evidence for auditory binding as information from all dimensions (both relevant and irrelevant) is compared across trials. However, evidence for binding also seems to interact with the nature of partial repetition. For example, Dyson (2010) developed auditory stimuli in which one aspect of the stimulus was relevant and the other aspect of the stimulus was irrelevant. Consequently, complete repetition could now be more appropriately compared to a

case of partial repetition in which only the relevant aspect of the stimulus was maintained and the irrelevant aspect of the stimulus changed. Also, complete change could now be more appropriately compared to a case of partial repetition in which only the irrelevant aspect of the stimulus was maintained and the relevant aspect of the stimulus changed. In comparing such cases, processing difficulties associated with partial repetition were more apparent when the relevant dimensional value repeated relative to when the relevant dimensional value alternated. These finding suggests an important caveat in feature binding, in that binding is not necessarily mandatory but rather more likely under conditions of low processing demands, such as cases in which the relevant aspect of the stimulus is maintained from trial $n-1$ to trial n . In addition to broad theoretical correspondence between Dyson (2010) and Zmigrod and Hommel (2008), the data also revealed patterns unique to audition, in that both studies reported that when stimuli were composed of location and pitch information, location was more influenced by irrelevant variation of pitch than vice versa. Moreover, Zmigrod and Hommel (2008; Experiment 1) found that loudness was more influenced by irrelevant variation in pitch than was pitch by irrelevant variation in loudness. If pitch must in fact be attended to in order to bind auditory features (as per FB-FIT; Woods et al., 2001), then an explanation of the above asymmetries becomes clear, namely that pitch receives preferential processing leading to difficulty in ignoring this particular auditory feature. Thus, when features such as loudness and location are coupled with pitch, pitch will always be attended to, leading to greater failures of selective attention when attending to location and loudness, compared with attending to pitch. Despite the asymmetries observed in auditory feature integration, these studies demonstrate that auditory features are bound in a similar manner to visual features. Moreover, object file and event file theories appear to be extendable to the auditory domain with minor allowances being made in terms of the critical

feature across which other stimulus information from that modality may be bound (e.g., location in vision, pitch in audition).

Intermodal feature binding

Given that there is evidence to suggest that event file theory is broadly equivalent across vision and audition, Zmigrod, Spape and Hommel (2009), conducted a study to determine if such an account could be extended to the multi-modal domain. The authors used audio-visual stimuli, in which the audio component consisted of 1000 or 3000 Hz pure tones and the visual stimuli consisted of a red or green circle, and asked participants to selectively attend to, and classify the stimuli based on either the colour (in the block in which vision was task-relevant) or the pitch (in the block in which audition was task-relevant). In addition to determining if interaction patterns similar to those observed in studies examining repetition effects within a modality would be seen for multimodal stimuli, the authors were also interested in determining if specific effects of modality would be observed. In terms of the first objective, the data provided support for the application of event file theory to multimodal stimuli. They found that, as with objects composed of within-modality features, partial repetition lead to a slowing of responses for objects composed of between-modality features. This result suggests that multimodal stimuli are also spontaneously integrated into a unified percept. As with intra-modal integration, feature binding occurred despite the fact that participants were instructed to focus on only one modality (the task-relevant modality), once again supporting the analysis of irrelevant variation as an index of binding. In terms of modality, the data also suggest that participants were faster at responding to the visual modality than the auditory modality across all inter-trial contingencies. Unfortunately, the authors did not attempt to obtain baseline measures for speeds of processing for the different signals, as such, the asymmetry may not be central to the architecture of audio-

visual binding but rather due to the fact that the visual signal was processed faster than the auditory signal. Nevertheless, the results clearly support the notion that modality is treated analogously to a feature (see also Benjamin, van der Smagt & Verstraten, 2008). In other words, regardless of whether the features comprising the stimulus are from the same or from different modalities, patterns of responses reflecting feature binding appear to be similar (pending certain spatio-temporal constraints; Koelewijn, Bronkhorst & Theeuwes, 2010). While the integration of auditory and visual features in this study revealed an asymmetry in which vision was less affected by audition than vice versa, such asymmetries also occur when binding features within a modality. However, without attempting to equate for speed of processing between modalities, the significance of this finding is ambiguous. Most importantly, the data provide strong support for a template account of audio-visual integration in which both relevant and irrelevant aspects of an object are available, such that irrelevant variation (partial repetition) across trial disrupts object / event file access.

Further evidence for audio-visual binding is reflected in a general finding in behavioural studies on intermodal or multimodal integration (MMI). Specifically, RTs are faster to bimodal stimuli than to either of the uni-modal components comprising the bimodal stimuli. This effect, known as the bimodal advantage (Miller, 1982), has been reliably demonstrated in behavioural studies on MMI. Giard and Peronnet (1999) conducted a study in which participants were asked to classify objects (uni-modal and bimodal in separate blocks) based on visual features (horizontal or vertical ellipses) or auditory features (high or low pure tones). Their results showed that participants were significantly faster at classifying the bimodal stimulus (regardless of which modality classification was based on) than the uni-modal stimulus. Molholm, Martinez, Shpaner, & Foxe (2007) conducted a similar study; however they used realistic objects

in contrast with the simple (abstract) objects used in the study by Giard and Peronnet (1999). Their results revealed that the findings of Giard and Peronnet, namely a bimodal advantage, could be extended to the classification of ‘complex’ audio-visual (AV) objects. In addition to behavioural evidence, event related potential (ERP) studies also highlighted the neural correlates underlying bimodal processing. Typically, the processing unique to bimodal processing is assessed by subtracting the sum of activations to the uni-modal features from the activation to the bimodal stimuli [$AV - (A + V)$] (Barth, Goldberg, & Brett, 1995; although see Teder-Salejani, Di Russo & Hillyard, 2002 to an objection to this formula). Any difference is taken to index additional processing of the bimodal stimuli that is not present for the uni-modal components. Stated differently, a non-zero difference suggests that processing a bimodal object is not the equivalent of simply processing each of the uni-modal components comprising the bimodal object. Many ERP studies on multimodal integration have in fact reported relatively early influences when assuming this super additive model (Giard & Peronnet, 1999; Molholm et al., 2007; Talsma, Doty & Woldorff, 2006) with bimodal effects being observed 50 to 100 ms post stimulus. The early appearance of enhanced processing in ERP studies, in addition to faster reaction time to multimodal stimuli seen in behavioral studies suggests that audio-visual binding is likely pre-attentive.

Factors affecting MMI

Although research into bimodal perception consistently demonstrates both behavioral and electrophysiological advantages of processing bimodal compared to uni-modal stimuli, the specific patterns of responses to the different sensory signals of the bimodal stimuli provide less consistent results. Overall, this body of research demonstrates the possibility of an asymmetry in

processing bimodal signals, in that one modality may influence registration of the other modality to a greater degree (see previous discussion of Zmigrod et al., 2009). These differential patterns of results obtained across various studies of multimodal integration suggest that multimodal integration is not a rigid process, but is instead is a dynamic process that adapts based on the situation and the individual. If modalities are like dimensions (Benjamin et al., 2008) then it is possible to make claims about the architecture of audio-visual binding in the same way it has been possible to argue that location is to vision as pitch is to audition during intramodal binding (c.f., Kubovy, 1988; Wolfe, 1994; Woods et al., 2001). For example, is there any evidence to suggest that visual information dominates auditory information (or vice versa) during the binding of auditory and visual information? Before attributing such differences to architectural differences, it is important to consider the contribution that methodology has upon the observation of different modality biases. Specifically, the stimuli, task requirements, attention load and expertise employed in a study may influence the pattern of results obtained. Each of these factors and their potential influence on bimodal perceptual tasks will be reviewed below.

Speed of processing

One potential explanation for modality asymmetry is unequal speeds of processing of the uni-modal components comprising the multimodal object (Ben-Artzi & Marks, 1995; Melara & O'Brien, 1987; Patching & Quinlan, 2002; Pomerantz, Pristach, & Carson, 1989; Zmigrod & Hommel, 2009). If there exists a significant difference in processing times for uni-modal stimuli, then when uni-modal stimuli are combined there will be different processing speeds when responding to one aspect of the bimodal stimuli over the other (e.g., Mordkoff & Yantis, 1991). Therefore, evidence for modality dominance may arise simply as a result of that modality being processed more efficiently than the other. Many studies have addressed this potential explanation

by attempting to equate uni-modal speeds of processing before combining these components to create the multimodal object. To this end, Ben-Atzi and Marks (1995) used various stimuli combinations in order to equate performance on responding to the visual or auditory components of the multimodal stimulus. Particularly, they used four pairs of tones (a high tone and a low tone) varying in the difference between the pitch of the two tones, and four pairs of visual dot locations (one above fixation and one below) varying in the difference between the vertical position of the dots. When they paired the stimulus such that multimodal classification based on either modality was equivalent, they found that participants were still faster at classifying the visual compared with the auditory signal during bi-modal processing. Their results suggest the discriminability and thus signal processing speed cannot provide a sufficient explanation for bimodal asymmetries on multimodal tasks. This finding was supported in a similar study by Patching and Quinlan (2002) who also found a persistence of visual dominance when baseline responding was equated. However in contrast to the findings of Ben-Atzi and Marks (1995) and Patching and Quinlan, (2002), a recent study by Yuval-Greenberg and Deouell (2009) demonstrated that equating the discriminability of visual and auditory features was in fact successful in eliminating visual dominance. In this study participants were presented with audio-visual (AV) stimuli consisting of realistic objects. A single trial consisted of an AV pairing followed by a word; the participants' task was to decide if the word represented the object they perceived in the task relevant modality (either A or V), which was assigned at the beginning of each block. For example, the participant may hear a cat's meow and an image of a dog, and would then be presented with the word 'Dog'. If participants were attending to the audio signal, they should respond that the word did not represent what they heard, whereas if they were attending to the visual signal they would respond that the word did in fact represent the image

they had seen. Also, on half the trials in each block the visual images were presented with high contrast and on half the trials they were presented with low contrast. This was successful in generating variation in speed of visual processing as the authors found that participants were significantly slower at responding when presented with the low contrast compared with the high contrast images. More importantly however, the data revealed that when participants were presented with high contrast images an asymmetry existed in which responses were faster when vision was the task relevant modality, compared to when audition was the task relevant modality. This asymmetry disappeared however, when low-contrast images were used, leading to equivalent response times when either vision or audition were deemed to be task-relevant. It should be noted that in the experiments by Ben-Artzi and Marks (1995) and Patching and Quinlan (2002) simple audio-visual stimuli were used, while the study by Yuval-Greenberg and Deouell (2009) employed realistic objects, which may in part account for the different results obtained.

Task Requirements

A second possible factor influencing multimodal performance may be the task employed in the research study. The notion of visual dominance is supported empirically by various illusions in which vision alters or biases perception in another modality, the most commonly cited being the ventriloquist effect (Radeau & Bertelson, 1974). In the ventriloquist effect, visual and auditory components of a stimulus are presented at different spatial locations resulting in the auditory signal being perceived as originating closer to the location of the visual stimulus. Another robust example of visual dominance is the Colavita Effect (Colavita, 1974). Colavita presented participants with uni-modal auditory and uni-modal visual stimuli on most trials (30 of 35), and asked them to press one button if the presented stimulus was visual (a flash) and another

button if the presented stimulus was auditory (a beep). However, on a small portion of trials (5 out of 35) auditory and visual stimuli were presented simultaneously. Participants were not given any specific instructions as to how to respond to the multimodal trials, and prior to the experiment, were in fact told that they were accidental occurrences. The results of this experiment showed that on 49 of the 50 multimodal trials (summed across all participants) participants responded using the visual response button. Furthermore, many participants reported not even hearing the sound.

While these examples show clear evidence for visual dominance, there have been various reports of auditory dominance (Ernst & Banks, 2002; Morein-Zamir, Soto-Faraco & Kingstone, 2003; Shams, Kamitani & Shimojo, 2002). In a study by Shams et al., (2002) tones and flashes were presented sequentially, with congruent presentation on half the trials (a single tone and flash, or, two tones and two flashes) and incongruent presentation on the other half of the trials (one tone and two flashes, or, vice versa). Participants were asked to report either the number of tones they heard or the number of flashes they saw. On incongruent trials, participants reported seeing the same numbers of flashes as there were tones; namely when a single flash was presented with two tones, participants reported seeing two flashes. This study was one of the first to demonstrate an instance in which auditory perception altered visual perception, hence providing support for auditory dominance.

Despite some general arguments that vision is the dominant sense for human perception in many situations (Posner, Nissen & Klein, 1976; Quinlan, 2000; Rock & Harris, 1967), it is clear that the nature of the task may influence the expression of modality dominance. For instance, visual perception putatively has high acuity for spatial resolution, while auditory perception putatively has high acuity for temporal resolution (Bermant & Welch 1976;

Bertelson, 1999; Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Recanzone, 2003). As such, there may be visual or auditory dominance based on whether the task requires a spatial or a temporal judgment respectively (see modality appropriate hypothesis, Welch & Warren, 1980) and the kinds of discrimination participants are asked to make. For example, in the ventriloquist effect the conflict between the visual and auditory stimuli is a spatial conflict, thus if vision is in fact the preferred perceptual system for spatial discrimination it is understandable that the visual stimulus would draw the auditory stimulus towards its location. Similarly, these types of argument also account for the instance of auditory dominance reported by Shams et al., (2002). In their experiment, flashes and tones were separated temporally, perhaps causing the perceptual system to rely most heavily on the auditory signal for this task, thus leading to illusory visual perception. Therefore, when considering audio-visual binding, it is important to respect the types of processing that seem most effective in both modalities.

Attentional load

Attentional load may also account for bimodal asymmetries. Posner et al., (1976) proposed that a further contribution towards visual dominance may be the weak alerting properties of the visual system. Specifically, he argued that humans must actively attend to visual stimuli, and unlike auditory or tactile stimuli which have strong alerting properties, visual stimuli can only alert if they are being attended to. Thus, Posner's account for visual dominance predicts that under high attention demands, humans defer to visual perception. If this account is correct, instances of visual dominance should be reduced or even reversed by reducing attentional task demands. Various studies have addressed the role of attention by manipulated attentional demands and observing the effects on bimodal asymmetries. In a study by Patching

and Quinlan (2002) visual dominance was present under conditions of divided attention (participants were required to attend to both modalities and were told following stimulus presentation which modality feature to respond to) but was significantly reduced under conditions of focused attention (participants were instructed as to which modality feature to respond to prior to stimulus presentation, and could thus focus attention on a single modality) , supporting the notion that higher task demands increase bimodal asymmetries. Similarly, a study by Sinnott, Spence and Soto-Faraco (2007) showed that the Colavita effect (bi-modal responding on the basis of visual rather than auditory values) could be modulated, although not reversed, by decreasing the attentional load of the task. Specifically, they presented participants with a higher number of uni-modal auditory than uni-modal visual trials, biasing attention towards audition, and were able to reduce the number of visual responses to the bimodal trials. It should be noted that in both of the studies reported above, visual dominance was reduced by attentional demands, but not completely eliminated or reversed. This suggests that while attention does influence multimodal integration, attentional accounts alone cannot explain bimodal asymmetries.

Expertise

Finally, perceptual expertise may also have the ability to bias modality dominance on multimodal tasks. There has been much research into visual expertise, suggesting that individuals who have visual expertise for a particular category of objects, show preferential processing of those objects compared with non-experts (Diamond & Carey, 1986; Tanaka, 2001; Tanaka & Gauthier, 1997; Tanaka & Taylor, 1991). Recently, there has also been research into so-called auditory experts, most often musicians. It has been demonstrated that musicians, compared with non-musicians, have better pitch discrimination abilities (Kishon-Rabin, Amir, Vexler & Zaltz, 2001), greater temporal resolution (Rammsayer & Altenmuller, 2006), and in the

case of conductors, better sound localization abilities (Munte, Nager, Beiss, Schroeder & Altenmuller, 2003; Nager, Kohlmetz, Altenmuller, Rodriguez-Fornells & Munte, 2003). While there has been research into visual and auditory expertise in isolation of each other, the effect of expertise (either visual or auditory) on multimodal tasks has not been directly investigated. However, in a recent study by Wong and Gauthier (2010), the researchers examined the impact of visual expertise on neural activity in non-visual areas of the brain. The results of their study revealed selectivity for musical notation (the object of the visual expertise) in widespread multimodal network areas when viewing musical notation compared with viewing roman numerals or mathematical symbols. They also reported that neural activity, in non-visual areas, to musical notation was able to predict musical expertise. While the task employed in the study by Wang and Gauthier (2010) was not multimodal, the results suggest that expertise in one modality affects processing in areas of the brain not specific to that modality (see Chartrand, Peretz & Belin, 2008, for a discussion of domain specificity in auditory expertise). Furthermore, in the study by Giard and Perronnet (1999), the authors analysed their data based on a post-hoc classification of participants as either auditory or visual dominant, based on participants being naturally faster at classifying uni-modal objects based on auditory features, or based on visual features respectively. The authors found differential patterns of neural activity among visual and auditory dominant participants during the processing of multimodal stimuli. Specifically, they reported that multimodal integration enhanced early processing almost exclusively in the non-dominant cortex of the participant. These results provide strong evidence for the online and dynamic nature of multimodal processing, suggesting that such processing adapts to the strengths of the individual by providing compensation to the non-dominant modality. While in this particular study expertise was based loosely on the modality that lead to faster classification on

uni-modal trials, these data coupled with the findings of Wang and Gauthier (2010), provide support for the notion that perceptual expertise in one modality could impact multimodal integration processes.

While there is clear evidence for bimodal asymmetries, with some instances of visual dominance and some of auditory dominance, only speculative accounts of the data are currently available. Bimodal asymmetries may arise due to sensory mechanisms, due to attentional mechanisms or due to both sensory and attentional mechanisms. A review of the multimodal literature highlights the importance of methodology, specifically decisions regarding the stimuli and task employed, the attentional load of the task, as well as the expertise of participants, when designing an experiment as these factors have the potential to direct the obtained data patterns.

Objectives of the current proposal

Many of the studies mentioned in the literature review provided thus far have used a modification of the standard Garner Paradigm (Garner, 1974; 1976) to investigate multimodal integration effects (Ben-Artzi & Marks, 1995, 1999b; Melara & O'Brien, 1987). These studies typically use stimuli comprised of two dimensions, *A* and *B*, from different modalities with each of the two dimensions containing two levels *A1*, *A2*, *B1* and *B2*. Performance on two speeded classification tasks, namely baseline and filter tasks, are compared to assess MMI effects. Specifically in a baseline task, the value of one (irrelevant) dimension is held constant while the value of the second (relevant) dimension is allowed to vary, and participants are asked to make a speeded classification of the value on the varying (relevant) dimension. In the filter task, all values are paired with all other values, and thus both the relevant and irrelevant values vary over trial, and participants are again asked to make a speeded classification of the value of a single

relevant dimension typically specified at the beginning of the block. Baseline and filter performance is then compared, with any increase in RTs and / or error rates in the filtering condition deemed to index interference of the irrelevant dimension (and hence binding). If irrelevant information cannot be suppressed as a result of binding between stimulus dimensions, then selective attention is said to have failed. While the paradigm described above compares performance across different tasks (baseline and filter), some researchers have compared performance across trial types in just the filter task, to investigate interference of the irrelevant dimension (e.g., Dyson, 2010; Hommel, 1998). Here, by considering the stimulus transitions between trial $n - 1$ and trial n in a filter condition, the value of the relevant dimension can either repeat (RR) or change (RC), and the value of the irrelevant dimension can either repeat (IR) or change (IC), leading to four possible inter-trial contingencies: RR-IR, RR-IC, RC-IR, RC-IC. Thus by comparing performance on RR-IR and RR-IC and between RC-IR and RC-IC trials, an estimate of the effect of irrelevance can be obtained. Analysis of the filtering condition alone has an advantage over the comparison of filtering and baseline conditions, as not only does the filter task require the suppression of the irrelevant dimension while the baseline task does not, but the filtering condition also contains more stimuli than the baseline condition. Therefore, it is not clear whether these performance differences arise from effects of irrelevant variation and / or from variation in stimulus uncertainty (see Dyson, & Quinlan, 2010, for a fuller discussion).

The first objective of the study was to ensure as far as possible that audio and visual processing was equated on an individual basis. A baseline type task was used to equate speeds of processing, however, rather than using a multimodal baseline, a uni-modal baseline task was used to obtain a purer measure of processing speeds for the auditory and visual stimuli. As noted above, the asymmetry observed in the study by Zmigrod et al., (2009) study may have been due

to different speeds of processing as the visual feature of the bimodal stimuli had a faster processing speed than the auditory feature. Thus, the current study attempted to match reaction times and error rates to the uni-modal components that comprised the bimodal stimuli. In addition to using a baseline measure to attempt to equate speeds of processing, the modality appropriateness hypothesis (Welch & Warren, 1980) was taken into account such that the task demands of the study were also selected to minimize bias towards one modality. Specifically, the study was designed such that the auditory response required a temporal discrimination and the visual response required a spatial discrimination, as these are putatively the strengths of these sensory systems (Bermant & Welch 1976; Bertelson, 1999; Bertelson & Aschersleben, 2003; Morein-Zamir et al., 2003; Recanzone, 2003). Thus, if speeds of processing and task appropriateness serve as a sufficient explanation for modality dominance, then in the current study, auditory and visual responses should be equivalent during both pre-training (discussed below) and uni-modal blocks.

A second objective of the study was to examine the inter-trial contingencies during filtering performance. Specifically by examining the magnitude of irrelevant variation on RTs and error rates during relevant repeat (RR-IR versus RR-IC) and relevant change (RC-IR versus RC-IC) trials, we examined the extent to which feature binding between modalities operates similar to feature binding within modalities. Zmigrod et al., (2009) provided initial support for equivalencies between dimensions and modalities, however, they did not directly interrogate the effects of task-relevance on stimulus-feature integration but rather on stimulus-response integration. We expected to see evidence for binding of audio and visual information indexed by an interaction between relevance and irrelevance, similar to studies examining intra-modal binding (Dyson, 2010; Dyson & Quinlan, 2004). Furthermore, based on these previous

studies we also expected that evidence of binding, as indexed by the influence of irrelevant variation, would be stronger during cases where the value of the relevant dimension repeated across consecutive trials. Furthermore, we were also interested in determining if an asymmetry would emerge based on the task relevant modality, bearing in mind that speeds of processing of the auditory and visual components comprising the stimuli should be equated, categorization was appropriate to each modality, and, the use of a selective attention task reduced the bias of visual dominance. Based on the design of the current experiment, we expected that there would be equivalent processing of the audio and visual signals prior to training (discussed below), and thus no evidence for modality dominance.

A third and final objective was to examine the modulation of AV binding as a function of experimentally-induced expertise. While the effects of perceptual expertise on perception in the modality of the expertise have been extensively studied, the effects of perceptual expertise on a multimodal task have not been systematically investigated. However, based on findings by Giard and Peronnet (1999) which revealed differential neural activations to a multimodal task based on participants' dominant modality, as well as Wang and Gauthier's (2010) data which revealed that visual expertise affected neural activity in brain regions not dedicated to processing visual objects, there is preliminary support for the notion that perceptual expertise in one modality may in fact influence multimodal integration processes. It should be noted that the few studies which have investigated the effects of expertise on multimodal task have been quasi-experimental, as expertise was obtained outside of the experimental setting. In the current study, we provided auditory or visual training in order to establish different perceptual 'expertise' in the laboratory. The training consisted of a special version of the filter task in which only one of modalities served as the task-relevant dimension. Feature integrations patterns before and after

training were compared to determine if the process of feature integration was affected by expertise. Results revealing differential patterns of feature integration before and after training, as well as across different training groups, would suggest that multimodal integration is not only affected by properties of the stimulus and the task, but also by the abilities of the individuals undertaking the task. Specifically, we expected that participants who received training in visual responding would show visual dominance post-training, and those who received training in auditory responding would show auditory dominance post-training.

Methods

Participants

A total of 37 participants completed the study. However, 5 participants were excluded due to an average error rate of greater than 20% across all experimental segments (uni-modal baseline, training, and filter blocks). Thus data from 32 participants was included in the analysis with 16 participants receiving auditory training and 16 receiving visual training. 66% of participants were female, with 11 females in auditory trained group and 10 in the visual trained group. The average age of participants was 27 years, (26 years in visual group and 29 in the auditory group). 84% of participants were right handed, with 14 right handed participants in the auditory group and 13 in the visual group. The participants were recruited from surrounding community and the undergraduate database (SONA) from Ryerson University and were offered course credits or money (\$20) in exchange for participation. All participants reported normal or corrected-to-normal vision and normal hearing.

Stimuli and Apparatus

Visual and auditory targets and cues were presented via PsyScope (Cohen et al., 1993), and participant responses were recorded using a PsyScope Button Box. Visual stimuli were displayed in the center of a 15 inch iiyama monitor, simultaneously with auditory stimuli which were played over Sennheiser HD202 headphones. Response times and error rates were recorded using a PsyScope Button Box. The experiment was conducted in a darkened, quiet room.

For the auditory stimuli, participants were asked to classify the object as A1 (a single auditory tone) or A2 (two auditory tones). A1 stimuli consisted of a 1000 Hz pure tone, with 5ms onset and offset ramps at the beginning and end of the sound. A2 stimuli were created from A1, by adding a 2 ms offset at 48 ms into the sound and 2 ms onset ramp at 50 ms into the sound, creating a 4 ms gap between the two tones. Auditory stimuli were presented at four different intensity levels for A1 and A2 (58, 60, 62, and 64dB) at the start of the experiment. For the visual dimension participants were asked to classify the object as V1 (a single grey rectangle) or V2 (two grey rectangles). V1 stimuli had an on-screen dimension of 90 x 60 mm. V2 stimuli were created from the V1, by placing a black (0 % luminance) 2mm vertical line down the center of the rectangle, creating two rectangles with the dimensions of approximately 44 x 60 mm. The background color against which all visual stimuli were presented was black with 0% luminance. Visual stimuli were presented at four different luminance levels for V1 and V2 (25, 28, 31, 34% contrast) at the start of the experiment. Both auditory and visual stimuli were presented for 100 ms.

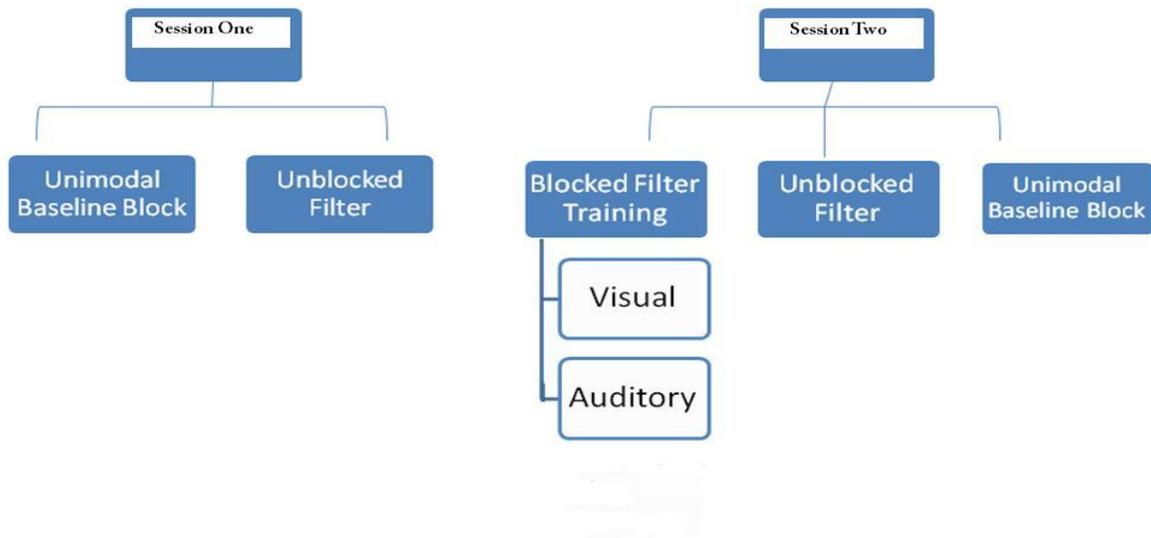
In each stage of the experiment, target stimuli were preceded by an audio-visual cue and followed by audio-visual feedback. The visual cue consisted of a printed letter A (for auditory) or V (for visual) presented in white in the center of the screen. The auditory cue consisted of a

spoken letter A (for auditory) or V (for visual), presented binaurally over headphones. The auditory and visual cues were presented simultaneously for 500 ms. The visual feedback consisted of a cross presented in the center of the screen. The cross was green for correct trials and red for incorrect trials. Auditory feedback in the form of a ‘ding’ and a ‘distorted ding’ was also presented binaurally over headphones, the former always following correct trials and the latter always following incorrect trials. The auditory and visual feedback were presented simultaneously for 500 ms.

Design

Each participant completed five tasks over two experimental sessions. During the first session, participants completed a *Uni-modal Baseline* task followed a *Bimodal Unblocked Filter* task. In the second session, participants completed a *Bimodal Blocked Filter Training* followed by a *Bimodal Unblocked Filter* task, followed by a second *Uni-modal Baseline* task (see Figure 1 for a schematic representation of the experimental design). In all tasks, participants were asked to make a speeded classification of value of the task relevant dimension, which was cued at every trial.

Figure 1. Schematic representation of experimental design. Left side indicates the tasks participants take part in during the first session. Right side indicates the tasks participants take part in during the second session.



During Uni-modal Baseline tasks stimuli were presented uni-modally. Using the 4 intensity levels for A1 and A2, and the 4 luminance levels for V1 and V2, there were 16 possible stimuli in the uni-modal baseline task. Each uni-modal baseline block commenced with 24 practice trials taken randomly from the experimental trials, followed by 10 presentation of each stimulus type for a total of 160 experimental trials per block. For each participant, the Uni-modal Baseline task in the first session was analyzed immediately following participant completion, to determine the intensity level of the auditory stimuli and the contrast level of the visual stimuli that lead to the greatest correspondence between RTs and error rates for auditory and visual responding. The intensity and contrast level chosen for each participant were then used in the upcoming blocked and unblocked filter tasks.

During the *unblocked filter* stage, the task relevant dimension (auditory (A) or visual (V)), as well as the auditory (A1 or A2) and visual (V1 or V2) values were allowed to vary from

trial $n-1$ to n . There were four possible stimuli in the filter conditions, based on all possible orthogonal combinations of the two values on each dimensions, namely, $A1V1$, $A1V2$, $A2V1$ or $A2V2$. There were 16 possible inter-trial contingencies in the filter blocks, based on all possible orthogonal combinations of the relevant dimension on trial $n-1$ to n (i.e., $A-A$, $A-V$, $V-V$, $V-A$), and all possible combinations of the values of the relevant dimension (i.e., RR for relevant repeat trials or RC for relevant change trials) and irrelevant dimensions (i.e., IR for irrelevant repeat trials or IC for irrelevant change trials) on trial $n-1$ to n (i.e., $RR-IR$, $RR-IC$, $RC-IR$, $RC-IC$). However, only trials in which the relevant dimension repeated ($A-A$, $V-V$) were analyzed in the current analysis. Trials in each block were randomized per participant and all inter-trial contingencies appropriate to each condition was presented with equal probability, and consecutive trial relations were equally represented by all stimulus pairings. The filter stage commenced with 24 practice trial pairs selected at random from the experimental trials, followed by 5 blocks of 64 experimental trial pairs (4 stimuli x 16 inter-trial contingencies).

The *blocked filter training* was similar to *unblocked filter training*, however the relevant dimension was essentially fixed throughout. Specifically, in *blocked filter training* the participant was asked to respond to one dimension (either auditory or visual) for approximately 85% of the trial pairs with the task-relevant dimension remaining the same from trial $n-1$ to n . On the remaining 15% of the trial pairs – the *catch trials*- the task-relevant dimension would represent the non-trained dimension for some or all of the trial pair. For example, a participant that received auditory training was presented with 85% $A-A$ trials, and, 15% of trials distributed equally across $A-V$, $V-A$ and $V-V$ contingencies. The catch trials ensured that participants did not completely prevent perception of the irrelevant dimension (i.e., closing eyes on auditory training or removing headphones on visual training), ensuring all stimuli was perceived as

multimodal and that participants maintained focus and actively engaged in the task. As in the *unblocked filter* stage the values of the relevant and irrelevant dimension varied from trial $n-1$ to n , and all inter-trial contingencies (*RR-IR, RR-IC, RC-IR or RC-IR*) were presented with equal probability. The *blocked filter* stage commenced with 24 practice trial pairs, followed by 2 experimental blocks each of which was comprised of 76 trial pairs (64 experimental trials and 12 catch trials).

To assess the effects of experimentally-induced training, participants then completed another unblocked filter condition and another uni-modal baseline condition (parameters identical to pre-training).

Procedures

Participants were tested individually, over two sessions which lasted for approximately one hour each. Upon arrival participants were provided with a brief overview of the study before providing written consent for participation. Participants were seated in front of the computer monitor (at a distance of 57 cm) and fitted with headphones. Each trial commenced with a black screen presented for 250 ms, followed by a 500 ms audiovisual cue to indicate the upcoming task-relevant modality, and where appropriate followed by the presentation of auditory (via headphones) and / or visual (center of computer screen) stimuli where appropriate for 100 ms. Participants were asked to respond as accurately and as quickly as possible by pressing one of two response buttons on the PsyScope Button Box. For half the participants the left-most button represented a response to either A1 or V1, and the right-most button represented a response to either A2 or V2. This response mapping was reversed for the other half of the participants.

Participant responses were followed by a short 50 ms response-checking period followed by audiovisual feedback on the accuracy of the response for a duration of 500 ms. The next trial began with the presentation of a black screen, establishing a response-stimulus interval of 1300 ms. Participants were offered a short break between consecutive blocks.

Results

Data Exclusion

Error rates were included if the response at trial n-1 was correct but the response at trial n was incorrect. Individual trials RTs were excluded for each participant based on the participants mean reaction time in a given condition. Individual trials were excluded from data analysis if a participants RT in a particular condition deviated from the mean of that condition by more than 2.5 standard deviations.

Uni-modal Baseline

RTs and error rates (arcsine transformed; Winer, 1962) were analyzed to check for equivalence between uni-modal auditory and visual responding. RT's and error rates were analyzed only for the specific intensity levels selected for each participant. RT and error rate data were independently submitted to a 2 (training: auditory (A), visual (V)) x 2 (block: pre-training, post-training) x 2 (stimulus modality: auditory (A), visual (V)) repeated-measures Analysis Of Variance (ANOVA). While the groups were not differentiated by training at this point in the experiment, training was included as a factor to test whether there were differences in the groups prior to the experimental manipulation. Tukey's HSD test ($p < .05$) was used to follow up on significant interactions. See Table 1 for a full summary of the ANOVA results.

Table 1. Uni-baseline task reaction times and error rates ANOVA table.

Source	RT Analysis			Error rate Analysis		
	F	MSE	p	F	MSE	p
G	0.33	18 800	.567	0.18	.03	.671
UB	27.46	502 000	<.001	1.26	.16	.271
M	3.85	19 200	.059	2.19	.10	.150
G x UB	0.10	1864	.752	0.23	.03	.634
G x M	1.26	6303	.270	0.20	.01	.656
M x UB	0.48	1546	.495	0.52	.01	.476
G x UB x M	0.46	1501	.501	0.01	.00	.914

Note: All degrees of freedom = 1,30. ANOVA= analysis of variance. G= training group (auditory, visual). UB = uni-baseline block (one, two). M = modality (auditory, visual). Significant ANOVA terms in bold.

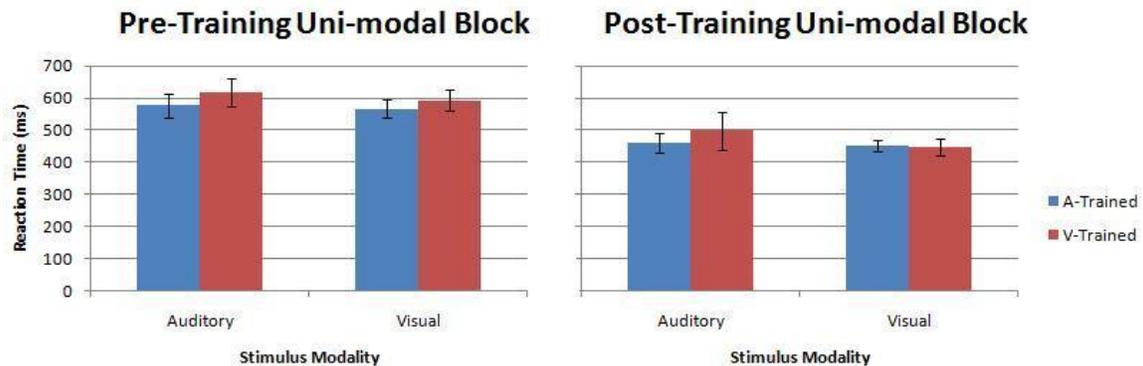
RT's

Reaction time analysis revealed no significant effects of stimulus modality, indicating that the speeds of processing for the auditory and visual signals were equated across participants ($A = 538$ ms, $V = 514$ ms; $p > .05$). There was also no significant effect of group, suggesting an absence of differences between the groups prior to training ($p > .05$). There was a main effect of block, indicating that participants were faster at responding to both modalities following training (pre-training = 596 ms, post-training = 459 ms; $p < .001$). The fact that block did not interact with group, ($p > .05$; see Figure 2), suggests that training had a nonspecific effect on uni-modal response speed, in that participants demonstrated an equivalent speeding to processing both modalities, rather than greater speeding to the trained modality. All other main effects and interactions were non-significant (all p 's $> .059$).

Figure 2. Reaction times for uni-modal blocks by training group pre and post-training.

Results suggest that overall participants are faster following training. Both pre-training and post-training there is no significant difference in auditory and visual responding, and no interaction

with training group. A-Trained=Audio Trained group. V-Trained=Visual Trained group. Error bars represent standard error.



Error Rates

There were no significant main effects or interactions in the error rate data (all p 's > .150). These results suggest that as with speeds of processing, uni-modal error rates were successfully equated for the auditory and visual stimuli, and did not differ by training group. However, unlike the RT results, there was no significant effect of block in the error rate data. This suggests that there is no speed accuracy trade off, in that, although participants were faster post-training, the error rates did not differ significantly between the two blocks, albeit with a trend for lower error rates following training (pre-training = 11.33%, post-training = 8.87%).

Unblocked Filter

Reaction time and error rate data (arcsine transformed; Winer, 1962) were independently submitted to a 2 (training: auditory (A), visual (V)) x 2 (block: pre-training, post-training) x 2 (stimulus modality: auditory (A), visual (V)) x 2 (relevance: repeat (RR), change (RC)) x 2 (irrelevance; repeat (IR), change (IC)) repeated measures ANOVA. Tukey's HSD test ($p < .05$) was used to follow up on significant interactions. See Table 2 for a summary of the 5 way repeated measures ANOVA for both RT and error rate data.

Table 2. Filter task reaction times and error rates ANOVA table.

Source	RT Analysis			Error rate Analysis		
	F	MSE	p	F	MSE	p
G	0.02	4029	.889	0.03	.01	.874
FB	24.42	576000	<.001	0.70	.04	.410
M	0.59	901	.718	0.01	.01	.924
R	17.75	130000	<.001	12.80	1.08	.001
I	7.03	15300	.013	0.34	.10	.565
G x FB	0.59	14000	.447	0.32	.02	.576
G x M	0.13	852	.726	0.52	.04	.476
G x R	1.17	8526	.288	0.99	.08	.328
G x I	0.09	199	.764	5.15	.15	.031
FB x M	0.68	4749	.415	<0.01	<.01	.971
FB x R	1.08	3845	.307	<0.01	<.01	.952
FB x I	0.01	24	.904	1.41	.05	.245
M x R	3.50	11600	.071	4.58	.22	.041
M x I	0.21	573	.649	0.88	.03	.355
R x I	2.86	18500	.101	12.41	.32	.001
G x FB x M	2.68	18600	.112	2.64	.11	.115
G x M x R	0.43	1424	.515	1.08	.05	.308
G x R x I	2.55	16400	.121	0.07	.00	.794
FB x M x R	1.16	5251	.290	0.82	.04	.372
FB x R x I	1.66	9791	.207	0.06	.00	.804
FB x G x I	4.26	6987	.048	0.88	.03	.123
M x R x I	<0.01	2	.982	0.23	.01	.638
G x FB x R	1.29	4,576	.266	0.44	.02	.512
G x M x I	0.81	2186	.377	<0.01	<.01	.987
FB x M x I	.200	731	.660	0.03	.01	.865
GxFBxMxR	1.86	8432	.182	0.28	.02	.599
GxFBxMxI	1.06	3,940	.311	1.11	.04	.301
GxFBxRxI	2.04	12000	.164	0.41	.01	.529
GxMxRxI	3.48	14300	.072	<0.01	<.01	.945
FBxMxRxI	0.56	4990	.459	8.42	.27	.007
GxFxMxRxI	3.63	32000	.066	0.01	<.01	.944

Note: All Degrees of freedom = 1, 30. ANOVA = analysis of variance. G = training group. FB = filter block. M = modality. R= relevance. I = irrelevance. All significant effects are presented in bold.

RT 's

RT analysis revealed a significant main effect of block ($p < .001$), showing faster reaction times post-training relative to pre-training (456 ms and 523 ms, respectively). A main effect of relevance was also found ($p < .001$), with participants responding faster to trials in which the relevant value repeated compared to when it changed (RR: 473 and RC: 505 ms, respectively). Finally there was a main effect of irrelevance ($p = .013$), with faster RT's to irrelevant repeat than to irrelevant change (IR: 484 and IC: 495 ms, respectively) trials. All other main effects were non-significant (all $ps > .718$).

Block, training group, and irrelevance were involved in a three way interaction ($p = .048$; see Figure 3). Follow-up two-way ANOVAs with the factors of training group and irrelevance were performed separately for each block (see Table 3 for the full ANOVA results).

Table 3. Table of Follow-up ANOVAs for training group and irrelevance by filter block in reaction times

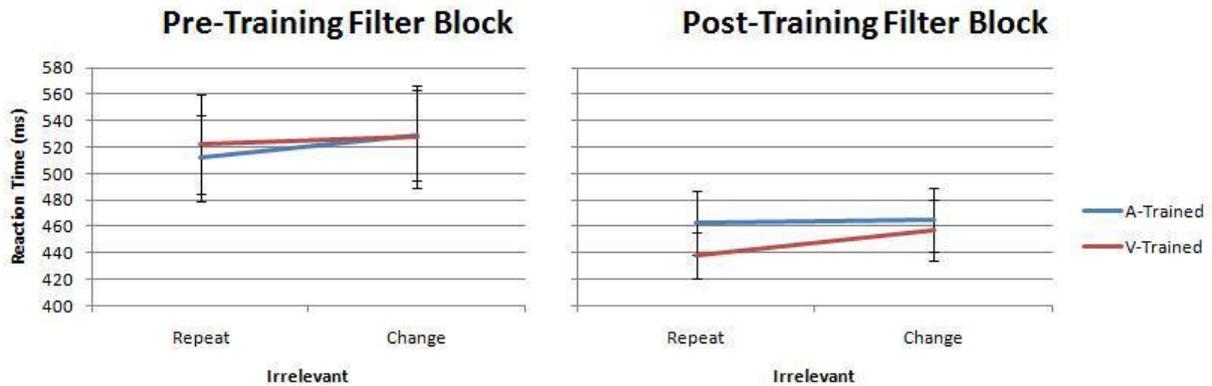
Source	Pre-Training Filter			Post Training Filter		
	F	MSE	p	F	MSE	p
G	.01	378	.924	.263	4136	.612
I	3.56	2071	.069	4.735	1766	.038
G x I	1.04	603	.317	3.2	1193	.084

Note: All Degrees of freedom = 1, 30. ANOVA = analysis of variance. G = training group. I = irrelevance.

The results revealed a significant main effect of irrelevance post-training ($p = .038$), however no significant interaction between training group and irrelevance prior to training ($p = .317$; see Figure 3) or following training ($p = .084$; see Figure 3). The significant effect of irrelevance post-training as opposed to pre-training seems to indicate increased AV binding as a function of practice. The three way interaction seemed to arise from significantly longer RT's to irrelevant change (IC) than to irrelevant repeat (IR) trials (457 and 438 ms, respectively) in the

visual group post-training, but no difference between irrelevant repeat and change trials (463 and 465 ms, respectively) in the auditory group post-training. These results suggest that while the groups were equally unaffected by irrelevant variation prior to training, following training, there was an effect of irrelevance with the visual group showing marginally greater interference from irrelevant variation than did the auditory group. However, this result should be interpreted with caution, as the two-way ANOVA between training group and irrelevance post training was non-significant.

Figure 3. Interaction between training group and irrelevance by Filter block (pre and post training). Post-hoc tests reveal no significant difference between any conditions. A-Trained =auditory trained. V-Trained = visual trained.



Error Rates

The error rate data revealed a significant main effect of relevance ($p < .001$), with lower error rates on relevant repeat (RR: 6.50 %) compared to relevant change (RC: 9.86 %) trials. This is consistent with the faster RTs observed above for RR relative to RC trials. All other main effects were non-significant (all $ps > .410$). Interactions between modality x relevance ($p = .041$), and, relevance x irrelevance ($p < .001$) were subsumed by a four way interaction between filter block, modality, relevance x irrelevance ($p < .001$; see Figure 4). Figure 4 shows a cross-over between relevance and irrelevance for visual stimuli pre-training, and, a cross-over between

relevance and irrelevance for auditory stimuli post-training. This provides some suggestion that visual information influenced auditory information (but not vice versa) pre-training but that auditory information influenced visual information (but not vice versa) post-training. This seems to be predominately represented by the separation between RR and RC trials, which appear pronounced for auditory stimuli relative to visual stimuli pre-training, but pronounced for visual stimuli relative to auditory stimuli post-training. Descriptively, the pre-training data suggest relevant auditory values were harder to switch between than visual values and that irrelevant visual values interfered with auditory processing to a greater extent than irrelevant auditory values interfered with visual processing. In contrast, the post-training data suggest relevant visual values were harder to switch between than auditory values and that irrelevant auditory values interfered with visual processing to a greater extent than irrelevant visual values interfered with auditory processing. In sum, the data reflect potential visual dominance pre-training but auditory dominance post-training.

Figure 4. Interaction between relevance and irrelevance, by modality and filter block. This interaction suggests that pre-training irrelevant visual information impacted auditory responding (RR-IR vs RC-IR trials), but irrelevant visual information does not impact visual responding. This effect reverses post-training.



Follow-up three-way ANOVAs on the factors modality, relevance and irrelevance were conducted for each filter block separately (see Table 4 for the full summary of the ANOVA). The main effect of relevance seen in the overall 5-way ANOVA was present both pre and post-training, indicating a tendency for participants to be faster when the relevant value repeated compared to when it changed (Pre-training filter RR = 6.80%, RC =10.39%; Post-training filter RR = 6.21%, RC = 9.34%).

Table 4. Table of follow-up three-way ANOVAs for modality, relevance and irrelevance by filter block in error rate data.

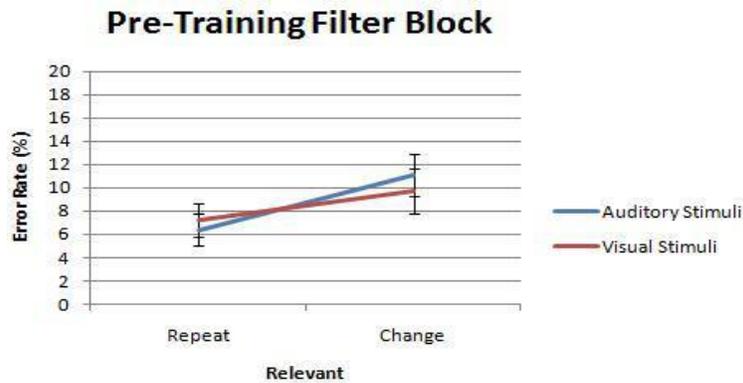
Source	Pre-Training Filter			Post Training Filter		
	F	MSE	p	F	MSE	p
M	.01	.00	.916	.00	.00	.959
R	10.99	.55	.002	7.66	.53	.009
I	2.03	.05	.164	.17	.01	.680
M x R	4.63	.23	.039	.70	.04	.409
M x I	.64	.02	.429	.23	.01	.632
R x I	6.91	.19	.013	5.55	.14	.025
M x R x I	3.20	.09	.084	5.24	.18	.029

Note: All degrees of freedom = 1. ANOVA= analysis of variance. M = modality (auditory, visual). R= relevance (repeat, change). I = irrelevance (repeat, change). Significant ANOVA terms in bold.

Pre-training data revealed no significant three way interaction ($p > .05$), however a two way interaction between modality x relevance ($p = .039$) was present (see Figure 5). This interaction was due to significantly greater errors to auditory stimuli when the relevant value changed (RC: 11.09 %) compared to when it repeated (RR: 6.41 %). For the visual stimuli error rates to relevant value repeat (RR: 7.19 %) and relevant value change (RC: 9.69 %) did not differ significantly (See Figure 5). These results confirm that response switching costs (RR versus RC trials) were greater for auditory than for visual stimuli pre-training, reflecting the ease of visual processing relative to auditory processing.

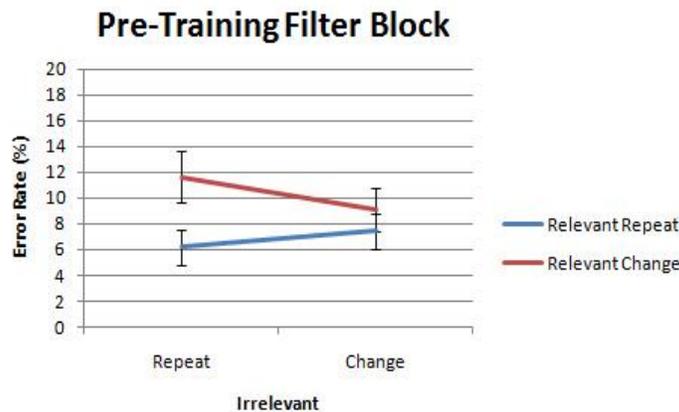
Figure 5. Interaction between modality and relevance in pre-training filter block.

Interaction shows that response switching was costly in auditory responding, as trials in which the relevant value repeated were significantly more errorful than trials in which the relevant value repeated. There was no significant difference between relevant repeat and change trials for visual responding.



A further two-way interaction between relevance and irrelevance ($p = .013$) was also shown. This interaction was due to a significantly greater error rate to relevant change compared to relevant repeat trials when the irrelevant value repeated (see left hand side of Figure 6; RR-IR: 6.17%, RC-IR: 11.64%), but not when it changed (see right hand side of Figure 6; RR-IC: 7.42%, RC-IC: 9.14 %). This pattern did not appear to be affected by modality.

Figure 6. Relevant repeat by relevant change interaction in pre-training filter block. This interaction suggests that participants were more accurate when both the relevant and irrelevant value repeated (RR-IR), compared with when the relevant value changed and the irrelevant value repeated (RC-IR)



Post-training data revealed a significant two-way interaction between relevance x irrelevance ($p = .025$) in addition to a three-way modality x relevance x irrelevance interaction ($p = .029$). The relevance by irrelevance interaction suggested that, as with pre-training, relevant change trials were more errorful than relevant repeat trials when the irrelevant value repeated but not when it changed. However, this relationship was mediated by modality (see right hand side of Figure 4). Thus follow up two-way ANOVA's for relevance and irrelevance were conducted separately for auditory and visual trials post-training (See Table 5 for a summary of the ANOVA). Specifically, visual responding revealed a significant interaction between relevance and irrelevance ($p = .002$) such that when the irrelevant value repeated participants were significantly less errorful on relevant repeat trials (RR-IR: 4.84 %), than on relevant change trials (RC-IR: 10.78 %). However when the irrelevant value changed, there was no significant difference between relevant repeat (RR-IC: 8.28 %) and relevant change trials (RC-IC: 7.19%). This confirms the observation that response switching was costly when processing visual stimuli post-training. For auditory responses, no significant interaction between relevance and irrelevance, ($p = .834$) was revealed, however, there was a main effect of relevance ($p = .018$), in that relevant repeat trials were less errorful (5.86%) than relevant change trials (9.69%). Therefore, post-training irrelevant auditory information influenced visual processing, but irrelevant visual information failed to have an influence on auditory processing. In this regard, the data provide some evidence for auditory dominance post-training.

Table 5. Table of follow-up ANOVAs on relevance and irrelevance by modality in post-training filter block in error rates.

Source	Auditory responses			Visual responses		
	F	MSE	p	F	MSE	p
R	6.30	.42	.018	2.77	.15	.106
I	.61	.02	.443	<.01	<.01	.987
R x I	.05	.00	.834	11.45	.32	.002

Note: All degrees of freedom = 1,30. ANOVA= analysis of variance. R= relevance (repeat, change). I = irrelevance (repeat, change). Significant ANOVA terms in bold.

Finally, the error rate data also revealed a significant training group x irrelevance interaction ($p = .031$), revealing a trend for the auditory group to be more accurate when responding to irrelevant change (IC: 3.03 %) compared with irrelevant repeat (IR: 2.78 %) trials, and the visual group showing the opposite pattern (IC: 2.61 %, IR: 3.05 %). However, post-hoc tests revealed no significant differences between any levels of these variables and so the interaction will not be discussed further.

Discussion

The results of this study reveal that participants were successfully matched in performance to uni-modal auditory and uni-modal visual stimuli. Furthermore, there is some support for the notion that modality is treated as a feature, indexed by effects of irrelevance, however, this effect was weak. Finally, despite successfully equating for uni-modal speeds of processing, as well as various additional steps taken to prevent modality asymmetries, audio-visual binding and response switching costs were greater pre-training for the auditory stimuli and greater post-training for the visual stimuli. These results indicate a trend towards visual dominance pre-training and a trend towards auditory dominance post-training. Training effects were general rather than specific in that both groups, auditory trained and visual trained, showed

improvement in both uni-modal and multimodal tasks following training. Moreover, the magnitude of the training effect, expressed as RT speeding, appeared to be equivalent between trained and untrained modalities. These main findings will now be discussed in greater detail.

Audio-visual binding

Evidence for audio-visual binding was present following training, indexed by irrelevance. That irrelevant information impacted responses to the relevant dimension suggests that even under explicit instructions to ignore this information, participants dedicated some processing resources to the irrelevant dimension. Generally this is taken to index a failure of selective attention, resulting from perceptual binding of the relevant and irrelevant information (e.g., Zmigrod, Spape & Hommel, 2009).

Audio-visual binding was also suggested in error rate data, by the interaction between relevance and irrelevance. Traditionally in the literature this interaction arises due to the fact that RR-IR trials are more accurate than RR-IC trials, and that RC-IC trials are more accurate than RC-IR trials (e.g., Hommel, 2004). This interaction may be explained by object file theory (Triesman & Gelade, 1980) in that any repetition from trial n to $n + 1$ causes the object file from trial n to be reused at trial $n + 1$. Thus partial repetition (RR-IC) is less accurate than complete repetition (RR-IR) due to the fact that in the former case, a no longer valid object file is recalled, evaluated and then rejected. For this same reason, namely recalling of an invalid object file due to partial repetition, RC-IR trials are also less accurate than RC-IC trials. The effects of irrelevance are generally stronger when the relevant value repeats (RR-IR and RR-IC) compared to when the relevant value changes (RC-IR and RC-IC). Resource allocation theory suggests that this asymmetry arises due to lower task demands when the relevant value repeats, thus leading to

unused processing resources being directed at the irrelevant dimension (Dyson, 2010). One of the primary objectives of the current experiment was to assess the fit of an object/ event file theory to multimodal stimuli. While Zmigrod et al., (2009) did use audio-visual stimuli to provide initial support for the idea that modality is treated like a feature, in that integrating features between modalities operate similar to integrating features across modalities, their study was focused on stimulus – response binding. Looking at 4 in the current study, suggests that the pattern of data typically found with intra-modal binding, is in fact present in the current data. Specifically pre-training auditory data show a positive slope when moving from RR-IR to RR-IC trials, indicating a trend to be more errorful on RR-IC trials. Moving from RC-IR to RC-IC trials shows a negative slope, indicating a trend to be more errorful on RC-IR trials. Compared with the auditory data, the visual data reveals relatively flat slopes, indicating no trend for a difference between any of the four trial types. Interestingly, this effect reverses when looking at the post-training data, as these trends are present for visual stimuli but not for auditory stimuli; this asymmetry will be discussed in detail below. While these traditional pattern of data can be seen with the current data, the contrasts between RR-IR and RR-IC and RC-IR and RC-IC were not significant either pre or post-training. Using irrelevant variation as an index of AV binding, these results suggest that any audio-visual binding effects in the current study were very weak compared to those found in previous studies (e.g., Ben-Artzi & Marks, 1995; Giard & Peronnet, 1999; Yuval-Greenberg & Deouell, 2009; Zmigrod, Spape & Hommel, 2009). However clear effects of irrelevance are present suggesting some level of audio-visual binding, providing support for the application of object/ event file theories to multimodal stimuli.

There are various potential accounts for the limited audio-visual binding shown in the current study. First, there is much evidence to suggest that audio-visual binding is strongest

when there is temporal as well as spatial coincidence (Meredith, Nemitz, & Stein, 1987). Because it was deemed important to create a task in which the audio response required a temporal discrimination and the visual response required a spatial discrimination, there was a temporal discrepancy in the stimuli used in the current study as the two auditory tones were presented sequentially, while the two visual rectangles were presented concurrently. This may have disrupted the temporal relationships between auditory and visual components in the present study. As one will recall, in the study by Shams and colleagues (2000), they used sequentially presented visual flashes and sequentially presented auditory tones, which lead to audio-visual binding. However, their task also promoted auditory dominance in that visual perception was influenced by the auditory stimuli as a result of the classification of stimuli according to temporal factors. Specifically, if the participants were presented with one auditory tone and two visual flashes, they reported only seeing one visual flash. This point highlights one of the challenges of studying multimodal stimuli, in that if a spatial discrimination or a temporal discrimination had been used for both modalities, then we would potentially bias the task towards one modality; however, using both temporal and spatial discriminations seems to have reduced the amount of cross-modal binding.

Another possible factor leading to the limited audio-visual binding may be that the visual stimuli were presented on the center of the computer screen, while the auditory stimuli were presented binaurally over headphones, possibly preventing the perception of spatial coincidence. However, given that previous studies have found audio-visual binding using headphones and computer monitors (e.g., Chen & Spence, 2009) this cannot provide a full account for the limited binding. Undoubtedly though, the use of free-field speakers positioned in the same depth plane as the computer monitor should facilitate audio-visual binding in future experiments.

It is worth noting that in a recent study, Patching and Quinlan (2002) found congruency effects in the absence of Garner interference. Specifically, the authors found that congruency effects (within trial effects) were present under both divided and selective attention conditions. However, they found that the effects of irrelevance (between trial effects) were only present under divided attention conditions and only for auditory responding. These results suggest that effects of irrelevance may not be necessary effects of binding, given congruency effects were present in the absence of effects of irrelevance. Pertinent to the current study is the fact that congruency in the absence of effects of irrelevance were seen more strongly (for both audio and visual responding) in the selective attention condition. Thus, the weak effect of irrelevant interference in the current study does not necessarily mean that the binding of the audio-visual stimuli was completely absent. Clarification on the extent of audio-visual binding in the current study could thus be provided by examining the data for effects of congruency. It would be expected that congruency would facilitate responding (i.e., one visual object and one auditory object), while incongruency would interfere with responding (i.e., one visual object and two auditory objects) if audio-visual binding was present. We might also expect an interaction between congruency and modality, indicating that congruency would facilitate responding to the non-dominant modality more than the dominant modality, and that incongruency would interfere with responding to the non-dominant modality more than to the dominate modality. The data have not yet been analyzed in terms of congruency effects, however this is an avenue that will be pursued, and that may provide additional evidence for audio-visual binding.

Modality Dominance

The data from the current experiment suggest that modality dominance was present, despite various steps taken to mitigate such effects. Specifically, at the level of stimulus, we successfully equated for uni-modal processing in terms of both error rates and RTs for the audio and visual dimensions of the bimodal stimuli. In terms of study design, we used a bimodal cue, rather than a visual only cue in order to prevent priming of the visual modality which may contribute to modality dominance. Furthermore, in terms of task demands, the literature suggests that under low compared with high task loads, modality dominance effects are significantly reduced (Sinnott, Spence & Soto-Faraco, 2007) and even eliminated completely (Yuval-Greenberg & Deouell, 2009). Thus, we chose a selective attention task over a divided attention task (essentially by placing the modality cue before rather than after the bimodal stimulus), to allow participants to focus their attention on the task relevant modality, rather than dividing attention across both modalities. Finally, at the level of participant expertise, we provided participants with training to establish either auditory or visual training in participants, to systematically examine the effects of expertise in one modality on a multimodal task, thus ensuring that an equal number of participants were auditory experts and visual experts.

As discussed above, there is some indication in the data that effects of irrelevance were present, indexed by partial repetition costs (RR-IC, RC-IR) against complete repetition (RR-IR) and complete alternation (RC-IC) facilitation. However, these effects were only suggested in the auditory stimuli pre-training and the visually stimuli post training. Furthermore, these effects were not mediated by training group. Partial repetition costs and complete repetition/alternation effects can only arise if information from both the task-relevant and the task-irrelevant modality are processed. While these individual effects were non-significant, the significant four-way

interaction (see Figure 4) provides some indication in the data that auditory responding was influenced by task-irrelevant visual stimuli prior to training to a greater extent than visual responding was influenced by task-irrelevant auditory stimuli, and vice versa in the post training data.

Further consistent evidence for modality dominance in the current experiment is provided by the response switching data. Previous research with perception-action coupling, specifically event-file theory, suggests that switching a response requires more processing than repeating a response (Hommel, 1998). Thus it is possible that if the perceptual component of a task is difficult and therefore requiring greater processing resources, fewer resources will be available for response selection. This would then lead to greater costs of switching a response compared to repeating a response. Conversely, if the perceptual component of the task was relatively easy, and therefore requiring less processing resources, greater resources would be left over for response selection, leading to a smaller cost of switching compared with repeating a response.

In the current experiment response switching was costly in terms of increased error rates to RC (response switch) compared with RR (response repeat) trials for the auditory stimuli. These effects were present both pre- and post-training, however, pre-training the effect was mediated by irrelevant visual information, in that response switching costs were only present for trials in which the irrelevant visual information repeated (RR-IR and RC-IR). Post training data revealed an overall effect of relevance for auditory stimuli (RR and RC), indicating that response switching was more costly than response repetition regardless of variation in the irrelevant visual information. Furthermore, the costs were greater prior to training (8%) than following training (4%), suggesting that training not only decreased the amount of audio-visual binding when

responding to auditory information, but also that the cost of switching responses decreased following training.

The visual data show a symmetric pattern to the auditory data in that response switching was more costly, and audio-visual binding was stronger post-training. Specifically, pre-training, there was neither a main effect of relevance nor any interaction between relevance and irrelevance, reflecting a lack of audio-visual binding. However, post-training, error rates to RC-IR (11%) compared to RR-IR (5%) trials were significantly greater, and were mediated by irrelevant auditory information, as the effect of response switching was only present when the irrelevant auditory information repeated. Thus, opposite to the auditory data, prior to training participants were able to selectively attend to the visual information and were efficient at switching between responses. However, post-training, participants displayed a failure of selective attention, and became less efficient at switching responses. What is critical here though is the lack of difference in RT or error rates for auditory and visual processing prior to bi-modal performance. Therefore, whilst controlling for standard notions of processing efficiency, the results can then be interpreted to suggest that the auditory task was more difficult than the visual task pre-training (suggesting visual dominance), and the visual task was more difficult following training (suggesting auditory dominance).

As mentioned above, one unexpected finding in the current experiment is that modality dominance was not affected by training group. It is not that surprising that pre-training, participants displayed visual dominance, as this is a common finding in the literature (e.g., Ben-Artzi & Marks, 1995; Giard & Peronnet, 1999; Yuval-Greenberg & Deouell, 2009; Zmigrod, Spape & Hommel, 2009), and some have even suggested that human perception is naturally visually dominant (Posner, Nissen & Klein, 1976). While it is surprising that training did not

influence modality dominance, an interesting aspect of the data is that participants displayed a trend for auditory dominance following training. This trend towards auditory dominance post-training may be reflective of the reduced task load following training. While the training did not have a specific effect on performance, training did in fact have the general effect of reducing the task load. This is indexed by the main effect of filter block in the reaction time data, which shows that all participants were more efficient (revealed by RT speeding) at the task following training. Previous studies on audio-visual integration have found that reducing task load effectively reduces visual dominance (e.g., Yuval-Greenberg & Deouell, 2009; Sinnott et al., 2007). Thus the elimination of visual dominance from pre to post-training may be due to a reduced task load following training due to more efficient processing of both the audio and the visual stimuli. One finding that is unique in the current study, is that the decreased task load leads to a trend towards auditory dominance. Previous studies which have decreased/eliminated visual dominance by manipulating task demands, generally find no modality asymmetry under manageable task demands. That the asymmetry reverses to favor auditory dominance in the current study may be reflective of the fact that speeds of processing for the auditory and visual signal were equated from the start. Specifically, previous studies may not have had the power to detect auditory dominance under manageable task loads due to unequal speeds of processing to begin with which bias the task towards visual dominance.

A further index of modality dominance in the current study may be offered by modality-switching data. Specifically from trial n to trial $n + 1$, participants can make consecutive auditory responses (A-A) or consecutive visual responses (V-V), or participants can make an auditory followed by a visual response (A-V) or a visual followed by an auditory response (V-A). Analyzing the data in such a manner, we would expect that pre-training V-V responding

would be greater than A-A responding, and furthermore, that A-V responding would be more efficient than V-A responding, if vision is the dominant modality. Post training we would expect the reverse pattern such that A-A responding would be more efficient than V-V responding and that V-A responding would be more efficient than A-V responding if audition is the dominant modality following training. Although the data has not been analysed in this manner currently, this is an avenue to explore in more detail.

Uni-modal Responding

The results of the current study indicate that speed and accuracy were equated for auditory and visual responding, as there was no main effect of stimulus modality in the pre-training uni-modal baseline block. Furthermore, there was no interaction between training group and stimulus modality, indicating that both groups were equated on speed and accuracy of responding to the uni-modal auditory and visual stimuli prior to training. These are important findings, as the modality asymmetries discussed cannot be interpreted based on differential speeds and accuracy in responding to the uni-modal components comprising the multimodal stimuli. Previous studies looking at selective attention have found processing asymmetries similar to those found in the current study (e.g., Ben-Artzi & Marks, 1995; Giard & Peronnet, 1999; Yuval-Greenberg & Deouell, 2009; Zmigrod, Spape & Hommel, 2009), however, they could not rule out a speed of processing account.

Comparing uni-modal baseline responding prior to and following training, revealed that the multimodal training task was successful in improving responding to the uni-modal auditory and visual stimuli, as indexed by speeded processing of approximately 134 ms. This finding is consistent with previous findings that show improvement on uni-modal responding following

training with multimodal stimuli (e.g., Beer & Watanabe, 2009). This training effect was general as both groups showed equivalent improvement on responding to both the audio and the visual components, suggesting that training with bimodal stimuli improved participants' responses to both components of the stimuli, even when the training task required attention to be predominately directed at a single modality. Thus, although only weak binding effects were observed in the current study, there was an extent to which both auditory and visual information were being considered since both forms of responding benefited from training with a single modality.

Future Research

In the current experiment, there were general effects of training, namely RT speeding, however, there were no specific effects of training indexed by a lack of an interaction with training group and stimulus modality post-training. The lack of a specific training effect in the current study may have been due to the minimal training provided to participants, or may reflect a broader immunity of audio-visual integration processes to expertise in a specific modality. To shed light on this question, future experiments could use a more extensive training protocol. Previous studies looking at the effects of training on uni-modal visual or auditory perceptual tasks have found improvement following training, however, in these studies the training spanned from 2-8 sessions (Seitz, Nanez, Holloway, Koyama & Watanabe, 2005; van Wassenhove & Nagarajan, 2007). This suggests that the brief nature of the training provided in the current study may not have been sufficient to reveal specific training effects.

Alternatively, a less taxing approach to examining the effects of expertise would be to recruit individual with perceptual expertise garnered outside of the laboratory setting. One particularly

interesting group of participants would be bird experts who are able to identify birds both by sound (bird songs) and sight. These individuals would provide insights into whether modality asymmetries are due to inherent properties of the human perceptual system, or are reflective of greater experience with visual information, rendering people generally visual experts. While there is much research into the effects of perceptual expertise in one modality on processing uni-modal stimuli, research into perceptual expertise in one modality on multimodal processes is limited and should be explored in more depth, as such research has the ability to shed light on mandatory versus flexible processes associated with multimodal integration.

The current study provides some support for the application of object/event file theories to multimodal integration. This support comes from the modest effects of irrelevance in a manner that is typical in studies of intra-modal binding. However, as noted earlier, in the current experiment, there were various reasons as to why audio-visual binding may have been weak. Thus, it is difficult to tell if traditional patterns of irrelevance were weak in the current study due to the fact that the features of the stimuli were from different modalities, or due to the fact that the experiment did not provide optimal conditions for audio-visual binding (done intentionally to reduce modality asymmetries). Thus future studies could use more typical audio-visual tasks in which binding is generally seen, and examine the effects of irrelevance under conditions in which previous research suggests that audio-visual integration should be strong. For example, rather than using different discriminations for the audio and visual stimuli based on the strength of each modality, studies could use a common discrimination for both modalities that is not a particular strength of either modality.

In the current study, although speeds of processing of the uni-modal auditory and visual stimuli were equated, trends for modality dominance were found both prior to and following

training. Accordingly a future study could attempt to equate speeds of processing using bimodal stimuli. In the current experiment, we provided evidence against an account of modality dominance based on uni-modal speeds of processing. Specifically, the data suggested that processing one modality of a bimodal stimulus is not equivalent to processing a uni-modal stimulus. Thus, one possibility for a future study would be to establish discrimination bimodally, and to present stimuli features asynchronously. For instance, if it was found that visual features were responded to faster, a future study could present visual features after auditory features allowing more time for auditory processing, evaluating the impact of extra processing time on reducing/reversing the asymmetry. Furthermore, presenting stimulus features asynchronously could provide insights into how much temporal asynchrony can be tolerated before bimodal feature integration does not take place (i.e., temporal window of integration).

General Conclusions

With respect to uni-modal responding the findings of the current experiment offer a successful method by which to equate uni-modal speeds and accuracy of processing for the component parts of a bimodal stimulus. The data also provide initial support for the notion that modality is treated analogously to a feature during cross-modal binding, given the trend for traditional effects of irrelevance. Furthermore, the data suggest that training with bimodal stimuli can in fact improve responding to uni-modal stimuli in terms of reaction times and error rates. The data also highlight the persistent nature of modality dominance, as this effect was present both prior to and following training, despite many controls to prevent this. The current data provide evidence against accounts of modality dominance based on issues such as speeds of processing, or the type of perceptual discrimination, as our stimuli were matched on these factors

and still produced evidence for modality dominance. Finally, the current experiment provides initial evidence that perceptual training with multimodal stimuli benefits performance on both subsequent uni-modal and multimodal tasks.

References

- Allport, D. A., Tipper, S. P., & Chmiel, N. (1985). Perceptual integration and post-categorical filtering. In M. I. Posner and O. S. M. Marin(eds). *Attention and Performance, 11*, 107-132.
- Barth, D.S., Goldberg, N., & Brett, B. S. (1995). The spatiotemporal organization of auditory, visual, and auditory–visual evoked potentials in rat cortex. *Brain Research, 678*, 177–190.
- Baumann, S., Meyer, M. & Ja'ncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience, 20*, 2238–2249.
- Ben-Artzi, E., & Marks, L., E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics, 57(8)*,1151-1162.
- Benjamins, J.S., van der Smagt, M.J. & Verstraten, F.A. (2008). Matching auditory and visual signals: is sensory modality just another feature? *Perception, 37*, 848–858.
- Bermant, R.I., & Welch, R.B. (1976). The effect of degree of visual-auditory stimulus separation and eye position upon the spatial interaction of vision and audition. *Perceptual and Motor Skills, 43*, 487–493.
- Bertelson, P. (1999). Ventriloquism: A case of cross-modal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347–362). Amsterdam: Elsevier.
- Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: cross-modal interaction on the time dimension: 1. Evidence from auditory–visual temporal order judgment. *International Journal of Psychophysiology, 50(1-2)*, 147-155.

- Besson, M., Schön, D., Moreno, S., Santos, A. & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.*, 25, 399–410.
- Chartrand, J. P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research*, 1220, 191-198.
- Colavita, F., B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409-412.
- Diamond, R., Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107-117.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4), 501-517.
- Duncan, J., & Humphreys, G., W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433-458.
- 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, 1170 - 1788
- Dyson, B. J., & Ishfaq, F. (2008). Auditory memory can be object based. *Psychonomic Bulletin and Review*, 15(2), 409-412.
- Dyson, B., J., & Quinlan, P., T. (2003). Feature and conjunction processing in the auditory modality. *Attention, Perception and Psychophysics*, 65(2), 254-272.
- Dyson, B. J., & Quinlan, P., T. (2004). Stimulus processing constraints in audition. *Journal of Experimental Psychology: Human Perception and Performance*, 30(6), 1117-1131.

- Dyson, B. J., & Quinlan, P. T. (2010). Decomposing the Garner interference paradigm: Evidence for dissociations between macrolevel and microlevel performance. *Attention, Perception, & Psychophysics*, *72*(6), 1676 - 1691.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-433.
- Garner, W. R. (1976). Interaction of stimulus dimension in concept and choice processes. *Cognitive Psychology*, *8*, 98 – 123.
- Girard, M. H., Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioural and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*, 473– 490.
- Hall, M. D., Pastore, R. E., Acker, B. E., & Huang, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception & Psychophysics*, *62*(6), 1243-1257.
- Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 163-169
- Hecht, L., Abbs, B., & Vecera, S. (2008). Auditory object-based attention. *Visual Cognition*, *16*(8), 1092-1147.
- Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*, 183–216.
- Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*, 494–500.
- Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research*, *71*, 42–63.

- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* 24(2), 175-219.
- Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *Journal of Basic Clinical Physiological Pharmacology*, 12(2), 125-43.
- Koelewijn T, Bronkhorst A, Theeuwes J (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychologica*, 134, 372–384.
- Melera, R. D., & O'Brien, T. P. (1990). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4), 323–336.
- Meredith, M.A., Nemitz, J.W., Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons: Temporal factors. *Journal of Neuroscience*, 7, 3215–3229.
- Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is multisensory: Co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, 26, 499–509.
- Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 66–79.
- Mordkoff, J.T., & Yantis, S. (1991). Dividing attention between colour and shape: Evidence for co-activation. *Perception and Psychophysics*, 53(4), 357-366.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, 17, 154–163

- Munte, T.F., Nager, W., Beiss, T., Schroeder, C., & Altenmuller, E. (2003). Specialization of the Specialized: Electrophysiological Investigations in Professional Musicians. *Annals of New York Academy of Science*, 999, 131-139.
- Nager, W., Kohlmetz, C., Altenmuller, E., Rodriguez-Fornells, A. & Munte, T.F. (2003). The fate of sounds in conductors' brains: an ERP study. *Cognitive Brain Research*, 17, 83–93.
- Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 755-775.
- Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1989). Attention and object perception. In B. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 53–89). Hillsdale, NJ: Erlbaum.
- Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information processing account of its origins and significance. *Psychological Review*, 83, 157-171.
- Rabbitt, P. M. A. (1992). Many happy repetitions: A celebration of the “Bertelson Repetition Effect” 1961–1991. In J. Alegria, D. Holender, J. Junca, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 313–330). Amsterdam: Elsevier.
- Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology*, 26, 63-71.
- Rammsayer, T., & Altenmuller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, 24, 37– 48.
- Recanzone, G.H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89, 1078–1093.

- Ro, T., Friggel, A., & Lavie, N. (2009). Musical expertise modulates the effects of visual perceptual load. *Attention, Perception & Psychophysics*, *71*(4), 671-674.
- Rock, I. & Harris, C. (1967). Vision and touch. *Scientific America*, *267*, 96-104.
- Seitz, A. R., Nanez, J. E., Holloway, S. R., Koyama, S., & Watanabe, T. (2005). Seeing what is not there shows the costs of perceptual learning. *Proceeding of the national Academy of Science*, *102*, 9080-9085.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*, 788.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*, 182-186.
- Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics*, *69*(5), 673-686.
- Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *Journal of Acoustical Society America*, *117*, 1285–1298.
- Talsma, D., Doty, T.J., and Woldorff, M.G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*, 679–690
- Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed audition objects. *Cognitive Brain Research*, *25*, 169–179.
- Tanaka, J. W. & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*, 43-47.
- Tanaka, J. W. & Gauthier, I. (1997). Mechanisms of Perceptual Learning. Goldstone, R. L.,

- Medin, D. L. & Schyns, P. G. (eds.) 83–125. Academic, San Diego, California.
- Tanka, K., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457-482
- Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., & Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *European Journal of Neuroscience*, *30*, 1636-1642
- Treisman, A. M. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.
- Treisman, A. M. (1999). Solutions to the binding problem: Review progress through controversy, summary and convergence. *Neuron*, *24*, 105–110.
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459–478.
- Treisman, A.M. & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- van Wassenhove, V., & Nagarajan, S. S. (2007). Auditory Cortical Plasticity in Learning to Discriminate Modulation Rate. *The Journal of Neuroscience*, *27* (10), 2663-2672.
- Welch, R. B., Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638-667.
- Wolfe, J. M., (1994). Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202–238
- Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2010). Conditions for face-like expertise with objects: Becoming a Ziggerin expert—but which type? *Psychological Science*.
- Woods, D. L., & Alain, C. (2001). Conjoining three auditory features: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *13*, 492–509.

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 65–74.

Yuval-Greenberg, S., & Deouell, L. (2009). The dog's meow: asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, 193, 4, 603-614.

Zmigrod, S., & Hommel, B. (2009). Auditory event files; Integrating auditory perception and action planning. *Attention, Perception & Psychophysics*, 71(2), 352-362.

Zmigrod, S., Spapé, M., & Hommel, B. (2009). Intermodal eventfiles: Integrating features across vision, audition, tactition, and action. *Psychological Research*, 73, 674–684.

