

NARROWING THE WINDOW:
MULTISENSORY PERCEPTUAL LEARNING AND ITS NEURAL CORRELATES

By

Albert R. Powers III

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

August 2010

Nashville, Tennessee

Approved by:

Calum Avison, Ph.D.

Randolph Blake, Ph.D.

Maureen Gannon, Ph.D.

Daniel Polley, Ph.D.

Mark Wallace, Ph.D.

*For my family
in both North and South.
I love you all.*

ACKNOWLEDGEMENTS

This work would not have been possible without the generous support of the Vanderbilt Medical Scientist Training Program, the Vanderbilt Kennedy Center, the National Institute of Deafness and Other Communication Disorders, and the National National Alliance for Research on Schizophrenia and Depression.

I would like to thank all those with whom I have had the pleasure of working throughout this and other projects. Dr. Andrea Hillock has been a partner with whom I have shared both the highs and lows of science, and her aid throughout my graduate career could not be more appreciated. Matthew Hevey has been invaluable in the completion of the work here and even more invaluable as a friend. I have the utmost faith that he will be an amazing physician. Leslie Dowell Kwakye has been there at the start of nearly every scientific idea I've had. I will consider myself blessed to call her and Gunnar my colleagues and friends throughout my career. Lastly, Zachary Barnett, Aaron Nidiffer, Vera Blau, Juliane Krueger, Matthew Fister, and Dipanwita Ghose have been wonderful colleagues and friends, and I know this is just the beginning of my work and friendship with them.

I am eternally grateful to Dr. Mark Wallace for his untiring support of my work, my professional growth, and my personal well-being. Outside of my own family, I have never encountered someone who has cared so deeply about my own success nor had so much faith in my abilities. At the conclusion of this process, I consider myself incomparably lucky to have gained such a wonderful life-long mentor and friend.

Lastly, I would like to thank my committee. To Dr. Avison, whose door has always been open to me, Dr. Blake, whose expertise has helped shape the way I think about my work, Dr. Polley, who has been the perfect role model for a young scientist, and Dr. Gannon, on whose constant mentorship I still rely: Thank you all so much for everything you've done.

TABLE OF CONTENTS

	Page
DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
Chapter	
I. INTRODUCTION	1
Multisensory Processes	1
Multisensory Integration in the Brain.....	3
Superior Colliculus and Anterior Ectosylvian Sulcus	4
Primate Cortical Areas	8
Superior Temporal Sulcus	10
Posterior Parietal Cortex: VIP and LIP.....	13
Frontal and Prefrontal Regions	16
Multisensory Integration in “Unisensory” Cortex.....	17
Principles of Multisensory Integration.....	22
Spatial Congruity.....	23
Stimulus Effectiveness	25
The Temporal Principle, The Multisensory Temporal Binding Window, and Cross-Modal Simultaneity Perception	26
Themes in Sensory Plasticity	32
Plasticity in Developing Animals.....	32
Plasticity in Adults.....	33
Multisensory Plasticity.....	36
Introduction to the Current Work.....	39
References	40
II. PERCEPTUAL TRAINING NARROWS	
THE WINDOW OF MULTISENSORY BINDING	69

Abstract	69
Introduction	70
Materials and Methods	72
2-AFC Training	72
Subjects	72
2-AFC Simultaneity Judgment Assessment	72
2-AFC Simultaneity Judgment Training	74
2-AFC Training Protocol	75
Follow-Up Assessment	75
2-AFC Exposure	75
Subjects	75
Exposure Protocol	76
2-AFC Exposure	76
2-IFC Training	77
Subjects	77
2-IFC Simultaneity Judgment Assessment	77
2-IFC Simultaneity Judgment Training	77
2-IFC Training Protocol	78
Follow-Up Assessment	78
Data Analysis	78
Grand Mean SOA Analysis	79
Window Size Estimation	79
Results	80
2-AFC	80
Judgments of audiovisual simultaneity can be used to define a multisensory temporal binding window	80
Perceptual training on a 2-AFC task results in a significant narrowing of the multisensory temporal binding window	82
Changes in the multisensory temporal window are not seen following passive exposure to the identical stimuli	83

Training-induced changes in the multisensory temporal binding window are stable for at least one week.....	85
2-IFC.....	87
Perceptual training on a 2-IFC simultaneity judgment task results in a significant narrowing in the size of the multisensory temporal binding window.....	87
Changes induced by perceptual training on a 2-IFC simultaneity judgment task are stable for at least one week.....	90
The window narrowing produced by 2-AFC and 2-IFC training tasks are highly similar in both degree of narrowing and its temporal dynamics.....	92
Large initial window size predicts success during training.....	94
Discussion.....	96
Acknowledgements.....	105
References.....	105

II. NEURAL CORRELATES OF MULTISENSORY PERCEPTUAL LEARNING ...111

Abstract.....	111
Introduction.....	112
Results.....	114
Perceptual training narrows the multisensory temporal binding window.....	114
Multisensory timing-dependent networks change with perceptual training ...	116
BOLD activity in pSTS decreases with training.....	120
Auditory and visual cortices as well as superior cerebellum exhibit decreases in BOLD activity after training.....	122
Resting state functional connectivity increases are seen between pSTS and auditory areas after training.....	125
Effective connectivity increases between pSTS and visual areas are seen after training.....	127

Discussion	130
Materials and Methods	135
Subjects.....	135
Procedure	135
Behavioral	135
Behavioral Assessment.....	135
Behavioral Training	137
Imaging.....	138
Procedure.....	138
Data Analysis.....	140
Estimation of Window Size.....	140
Imaging Data Preprocessing.....	141
Statistical Analysis.....	141
References	145
IV. GENERAL DISCUSSION	152
Summary of Results.....	152
Pertinent Themes.....	153
The Role of Unisensory Temporal Processing and Bayesian Integration.....	153
Multisensory Brain Networks, Connectivity, and Oscillations	157
Clinical Implications	160
Future Directions	164
Conclusions.....	172
References	173
APPENDIX: PERCEPTUAL TRAINING ON AN AUDIOVISUAL SIMULTANEITY JUDGMENT TASK ALTERS SOUND-INDUCED FLASH ILLUSION (SIFI) PERFORMANCE.....	183

LIST OF TABLES

Table	Page
3.1 Regions exhibiting localizer and pre-/post-training differences.....	119

LIST OF FIGURES

Figure	Page
1.1 Multisensory brain regions	9
3.1 Simultaneity Judgment Protocol	73
2.2 Training on a 2-AFC simultaneity judgment task narrows the temporal window of multisensory integration.....	81
2.3 Changes in the multisensory temporal window are not seen following passive exposure to the identical stimuli.	84
2.4 Training-induced changes in the multisensory temporal window are stable.....	86
2.5 Perceptual Training on a 2-IFC simultaneity judgment task significantly alters the size of the multisensory temporal binding window	89
2.6 Changes induced by perceptual training on a 2-IFC simultaneity judgment task are stable over time	91
2.7 The window narrowing produced by 2-AFC and 2-IFC training tasks are highly similar in dynamics and degree.....	93
2.8 Large initial window size predicts success during training.	95
2.9 Training without feedback fails to narrow the multisensory temporal binding window.....	99
2.10 Static participants on the 2-IFC task show increased response bias on some assessments.	101
3.1. Experimental procedure and behavioral results.....	115
3.2. Synchrony-responsive networks change with perceptual training.....	118
3.3. BOLD activity in posterior superior temporal sulcus (pSTS) decreases with training.....	121
3.4. Auditory and visual cortices as well as superior cerebellum exhibit decreases in BOLD activity after training.	123

3.5. Resting state functional connectivity increases are seen between pSTS and auditory areas after training.....	126
3.6 Effective connectivity changes between pSTS and visual areas are observed after training.....	128
4.1 Bimodal localization estimates and the ventriloquism effect	154
4.2 A proposed approach to modeling probability of audiovisual simultaneity judgment based upon unisensory temporal data	166

CHAPTER I

INTRODUCTION

“But the various senses incidentally perceive each other’s objects, not as so many separate senses, but as forming a single sense, when there is concurrent perception relating to the same object.” — Aristotle, De Anima

Multisensory Processes

While moving about in a world rife with sensory information, we are often blind to the fact that our senses work interdependently to construct a representation of the world around us. The enrichment we experience from the combination of these senses is seldom lost on us, however. Experiences as diverse as sipping a glass of wine to conversing in a crowded restaurant are enhanced by information from the different senses, and phenomena such as ventriloquism and even movies rely on the rules that govern their combination. In addition to these everyday occurrences, studies characterizing the improvements in comprehension of speech in noise with the addition of visual information (Sumbly and Pollack, 1954; MacLeod and Summerfield, 1987) and speeding of responses to multisensory stimuli (Miller, 1982, 1986) serve to quantify the benefits received when sensory stimuli are presented in tandem. The reason why these benefits arise from cross-modal integration is clear: because the individual senses provide information that is optimized to reveal specific characteristics of an object (i.e.,

spatial location for vision, timing for audition), the ability to combine that information synergistically maximizes what can be known about the object in question and about the physical world as a whole.

Illusions relying on conflicting contributions from the different senses have revealed a great deal about how their combination forms a singular and seamless percept. The ventriloquist effect, a form of religious expression and entertainment since ancient times (Connor, 2000), relies on the combination of visual motion from a dummy's mouth and concurrent sounds from the ventriloquist to create the illusion that the dummy itself is speaking. Similarly, the well-known McGurk effect relies on the combination of conflicting visual and auditory speech streams to produce a fusion percept that is completely distinct from the inputs of either two constituent streams (McGurk and MacDonald, 1976). Even when using highly reduced stimuli, information from one sense can be used to influence perception of information from another. This is the case with the sound-induced flash illusion (SIFI), wherein two brief tones presented in close temporal proximity to a single flash give rise to the illusory percept of two flashes (Shams et al., 2000, 2002). These phenomena are not limited to audiovisual interactions. For example, judgments of visual line orientation after centrifugal rotation are biased by altered vestibular input (Clark and Graybiel, 1965, 1966). A recent study has shown that puffs of air, timed to coincide with visual speech stimuli, are capable of altering the perception of the syllables spoken based upon whether or not those syllables contain a plosive consonant (Gick and Derrick, 2009). The use of auditory input for characterization of gustatory sensations was documented in a surprising study by Spence and colleagues, who found that the perceived crispness and freshness of potato chips was

highly influenced by the loudness and spectral weighting of the sounds they produce when chewed (Zampini and Spence, 2005). Similarly, alteration of the visual characteristics of food has been shown to affect the intensity of flavor perceived (DuBose et al., 2006), and even the time needed to reach satiety when eating (Rolls et al., 1982).

As these examples illustrate, our perceptions of the world are uniquely colored by the interactions between the individual sensory systems. However, the question as to where in the brain these interactions might occur is still a work in progress. The following section provides a brief outline of current knowledge concerning multisensory integration in the mammalian brain.

Multisensory Integration in the Brain

While the behavioral and perceptual benefits of combining information cross-modally have been known and described for thousands of years, our understanding of how and where the neural signals carrying that information might converge and be integrated in the brain is only in its nascent stages. The first characterization of multisensory integration focused upon how these processes are carried out on the level of a single multisensory neuron, here defined as one whose activity is influenced by stimulation from more than one sensory modality. In this case, multisensory integration occurs when the number of action potentials evoked by congruent information from different senses is significantly different from that evoked by the most effective unisensory stimulus when presented alone (Meredith and Stein, 1983). While the study of multisensory integration using single-unit electrophysiology is not the focus of this

volume, it is useful to consider integration in the deep layers of the superior colliculus as a model for how this process may be carried out in other brain areas.

Superior Colliculus and Anterior Ectosylvian Sulcus

The superior colliculus (SC) is a midbrain structure whose rich population of multisensory neurons (most commonly studied in the cat) makes it an optimal structure for the study of how integration may happen on the level of a single cell. The structure itself is made up of seven cellular and fibrous layers, usually functionally defined as superficial (layers I-III), and deep (layers IV-VII), wherein superficial layers are primarily visual and deeper layers may be responsive to stimuli from all three sensory modalities (Casagrande et al., 1972; Stein and Meredith, 1993). The anatomical convergence of sensory and motor inputs onto the deep layers of feline SC has been amply demonstrated (Huerta and Harting, 1984b, a), with principal visual inputs from the lateral suprasylvian and anterior ectosylvian areas as well as retina and lateral geniculate nucleus (Tortelly et al., 1980; Wallace et al., 1993), somatosensory inputs from the dorsal bank of the anterior ectosylvian sulcus along with substantial ascending contributions from the sensory trigeminal complex, dorsal column nuclei, lateral cervical nucleus, and spinal cord (Edwards et al., 1979; Stein et al., 1983; Huerta and Harting, 1984a), and auditory inputs from Field AES, the inferior colliculus and several other brainstem nuclei (Edwards et al., 1979; Meredith and Clemo, 1989). Importantly, input from these sources produces overlapping receptive field maps for the multisensory neurons upon which they converge, and this has allowed for classical quantitative analysis of multisensory interactions in the single cell.

As outlined above, integration may be said to have occurred when a cellular response to multisensory stimulation is significantly different from its response to the most effective unisensory stimulus (Meredith and Stein, 1983). This difference may take the form of response enhancement, wherein combined-modality stimuli result in a response profile that is greater than the response to the most effective unisensory stimulus, or response depression, wherein presentation of combined-modality stimuli results in a significantly smaller response than that evoked by the most effective unisensory stimulus. While the degree to which response depression has been described and tied to behavioral enhancements in SC and other multisensory structures has varied (Stein and Meredith, 1993), response enhancement has been a consistent hallmark of multisensory integration. The most dramatic examples of this enhancement are seen in the case of superadditivity, wherein presentation of a multisensory stimulus pair evokes a response that is greater than the combined responses to the individual unisensory events when taken alone (Meredith and Stein, 1983, 1986b; Meredith and Stein, 1986a).

Given that these multisensory interactions take place in cells that send efferent projections to centers in the spinal cord and midbrain and are responsible for generating orientation movements of the eye and head (Wallace et al., 1993), the behavioral relevance of these interactions is not difficult to surmise. If one of the functions of the SC is to detect and orient the animal to salient external events (Apter, 1946), and enhancement of activity in the descending neurons leading to orientation behaviors is observed with multimodal stimulus combination (Meredith and Stein, 1985), it stands to reason that presentation of spatially and temporally coincident stimuli from multiple modalities would aid detection of and orientation toward those stimuli. Indeed, this has

been observed in several studies in which awake, behaving cats are trained to orient to and approach auditory and visual stimuli (Stein et al., 1988; Stein et al., 1989), and the SC has been shown to be integral in orientation movements in humans (Leo et al., 2008). In a striking behavioral analogy to the superadditivity observed in single SC neurons, the number of correct responses (defined as direct approaches to the presented stimulus) to audiovisual combined stimuli was found to be greater than the sum of the number of correct responses to visual or auditory when presented alone (Stein et al., 1988). The behavioral importance of multisensory integration in the SC is similarly reinforced when considering the observation that the greatest degree of response enhancement is seen when two weakly effective unisensory stimuli are combined—the so-called principle of inverse effectiveness (Meredith and Stein, 1986a). In this construct, unisensory stimuli that themselves evoke minimal responses may produce enormous gains when presented in combination. This makes intuitive sense when considered in the context of the behavioral benefits conferred by multisensory integration: we stand to benefit most from the cross-modal combination of information when that information presented alone is least likely to be detected or accurately perceived. Not surprisingly, this principle has also been shown to be true in studies involving behaving cats (Stein et al., 1988), wherein the greatest accuracy gains are seen in the combination of weakly effective unisensory stimuli. This principle, along with the other rules for multisensory integration, is summarized in the next major section, *Principles of Multisensory Integration*.

While the electrophysiological and behavioral data described above, along with lesion studies producing profound visual hemispatial neglect (Casagrande et al., 1972) and a profound disruption in multisensory integration (Burnett et al., 2004) with ablation

of the SC indicate that this structure is remarkably important in conferring the perceptual and behavioral benefits of multisensory integration, the study of the neural correlates of multisensory integration has now been extended to a number of other brain structures.

One of the first cortical areas studied in detail from a multisensory perspective has been the anterior ectosylvian sulcus (AES) of the cat. Multisensory neurons in this structure seem to integrate inputs from the different senses according to the same rules as those that govern integration in the SC, with the same neurophysiological signatures of integration, and with the same co-registration of multisensory spatial maps as that seen in the SC (Wallace et al., 1992; Stein and Wallace, 1996). That the two structures appear to share a similar functional architecture makes intuitive sense, as multisensory integration in the cat SC is dependent upon active inputs from the AES: functional cooling of AES does not alter the responsiveness of SC multisensory neurons to multiple sensory inputs, but does strip them of their ability to integrate these inputs (Wallace and Stein, 1994). However, the mechanisms by which SC and AES neurons interact and the functional implications of those interactions are still unclear. While AES does contain multisensory neurons and does send projections to SC, these projections actually arise from unisensory neurons in the AES (Wallace et al., 1993). Moreover, while integration in SC is likely to confer benefits of improved detection and orientation to external events, it is unclear whether or not AES shares a similar function. This evidence taken together indicates that AES is a distinct—albeit not wholly independent—node of multisensory integration in the cat brain.

Primate Cortical Areas

The extension of multisensory electrophysiological study into cortical areas of the non-human primate has been slow, not least because a clear homologue to cat AES in primates has not yet been identified. However, it has long been thought that integration in the cortex must occur in association areas, following a classic feed-forward pattern of separate processing of the different streams followed by convergence. (This scheme has been challenged of late by evidence of integration in what had previously been thought of as ‘unisensory’ cortex; this will be discussed at length at the end of this section.) These association areas have been shown to receive corticocortical projections from primary sensory areas (Jones and Powell, 1970) and thalamocortical projections representing different sensory modalities (Burton and Jones, 1976). Additionally, these regions contain cells that are responsive to multiple modalities (Bruce et al., 1981), and even contain cells exhibiting auditory-visual interactions (Benevento et al., 1977). Given this evidence, it is no surprise that they emerged early as primary candidates for cortical multisensory integration (**Figure 1.1**). They include areas in and around the superior temporal sulcus (STS), areas in posterior parietal cortex (i.e., lateral intraparietal area [LIP], and ventral intraparietal area [VIP]), as well as areas within the frontal lobe (i.e., ventrolateral prefrontal cortex and ventral premotor cortex in the non-human primate). The characteristics of multisensory interactions in each of these areas are outlined below.

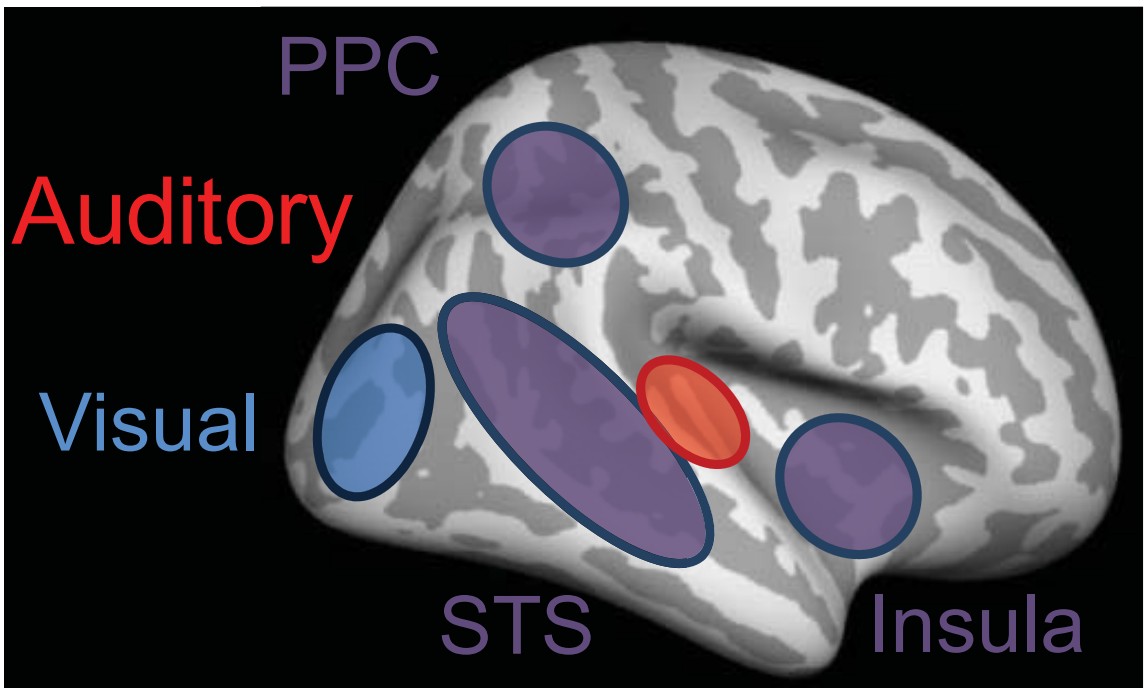


Figure 1.1. Multisensory brain regions. Regions of the brain in which multisensory interactions have been observed included traditionally-identified multisensory convergence regions (purple) such as superior temporal sulcus (STS), posterior parietal cortex (PPC), and insula, as well as visual (blue) and auditory (red) areas.

Superior Temporal Sulcus

Data from both electrophysiological studies in non-human primates and neuroimaging and evoked potential studies in humans indicate that STS is responsive to a wide range of auditory and visual stimuli (Beauchamp, 2005). Anatomically, STS shares abundant feed-forward and feedback connections with auditory belt regions in the superior temporal gyrus and the adjacent medial superior temporal area (MST) along with earlier visual areas (Barnes and Pandya, 1992; Hackett et al., 1998; Falchier et al., 2002; Cappe and Barone, 2005). Morphologically, the functionally-defined multisensory regions of interest within the sulcus itself most likely correspond to the cytoarchitecturally-defined area TPO in non-human primates (Seltzer and Pandya, 1994). Early electrophysiological studies estimate that 12-38% of neurons in STS are responsive to both auditory and visual stimuli, depending on the sub-region under scrutiny, with caudal regions of STS exhibiting the lowest percentages of audiovisual responsivity (Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988). Typically, the receptive fields of neurons in STS are extremely large; nearly all visually responsive cells sampled had receptive fields that extended into both visual hemifields, with a majority having receptive field sizes that approached the entirety of the visual field (Bruce et al., 1981; Hikosaka et al., 1988). Auditory receptive fields were similarly large and located mostly in the contralateral hemifield (Hikosaka et al., 1988). These studies indicate a lack of stimulus preference in visually-responsive neurons of the STS: no differential response was noted for visual stimulus size, shape, orientation, or contrast, with similarly no difference of effect with use of spots of light, slits, shadows, slides of complex objects, or three-dimensional objects (Bruce et al., 1981). However, a strong

directional preference for moving visual stimuli was noted in half of the neurons sampled (Hikosaka et al., 1988). Similarly, auditory responsive neurons did not show a preference for pure tone, white noise, human voice, or hand-clapping stimuli, and very few were selective for moving sound (Bruce et al., 1981; Hikosaka et al., 1988). From tracer (Seltzer and Pandya, 1994) and electrophysiological studies (Dahl et al., 2009) in non-human primates as well as high-resolution functional magnetic resonance imaging (fMRI) studies in humans (Beauchamp et al., 2004b), the structure of STS appears to be composed of a patchwork of unisensory auditory and visual-responsive neurons with bimodal patches interleaved. Given that projections to STS cortex from unisensory areas appear to be segregated into adjacent but non-overlapping patches (Seltzer and Pandya, 1994), these data suggest that input into STS is unisensory but is integrated via local corticocortical connections in multisensory patches.

Among the known multisensory convergence areas, none have been the subject of more intense study than STS. In particular, this structure has been shown to be important for the integration of complex, meaningful auditory-visual stimuli. For example, one electrophysiological study of the region has shown that 23% of neurons in primate STS that are responsive to biological motion (themselves 17% of the total number sampled) can be modulated by the addition of a congruent auditory stimulus (Barraclough et al., 2005). Results from similar human imaging studies indicate that middle and posterior sections of STS are most responsive to combined presentation of visual objects (tools, etc.) and their associated sounds (Beauchamp et al., 2004a; Stevenson and James, 2009), whereas the middle-anterior portion of STS has been shown to be preferentially responsive to audiovisual speech (Wright et al., 2003; Stevenson and James, 2009).

Interestingly, this region of STS has been shown to be important for the integration of letters and associated speech sounds (van Atteveldt et al., 2004), and integration deficits in adults with dyslexia are known to correlate with decreased activity in this area during a letter-speech sound matching task (Blau et al., 2009). Also in the realm of language, STS has been implicated as an area important in the comprehension benefits conferred by visual input in the presence of speech in noise (Ghazanfar et al., 2008; Bishop and Miller, 2009; Kayser and Logothetis, 2009), a finding that may well extend beyond mouth movements and into hand gestures as possible sources of congruent visual input (Hubbard et al., 2009). Strikingly, a recent study has demonstrated that focal inactivation of left posterior superior temporal sulcus (pSTS) by transcranial magnetic stimulation (TMS) is capable of disrupting the perception of the McGurk Effect (Beauchamp et al., 2010).

Perhaps most germane to this dissertation, STS also appears to be important in multisensory temporal processing. Several studies using both simple and complex stimuli in the context of both passive and active paradigms have implicated STS as being important for detection of audiovisual synchrony/asynchrony (Calvert et al., 2000; Bushara et al., 2001; Calvert, 2001; Macaluso et al., 2004; Noesselt et al., 2007). Interestingly, perception of the temporally-sensitive sound-induced flash illusion (SIFI), also appears to correlate with activity in posterior STS as well as primary visual cortices, indicating that perhaps STS could act as the conveyor of auditory information to visual cortices (Watkins et al., 2006; Watkins et al., 2007).

Posterior Parietal Cortex: VIP and LIP

The lateral intraparietal area (LIP) has been known for decades to be an important association area where information from different sensory streams is combined and commands are generated for attention to and exploration of peripersonal space (Mountcastle et al., 1975). Anatomically, LIP shares reciprocal cortico-cortical connections with various regions of extrastriate visual cortex, including parieto-occipital visual area PO, areas V3, V3A, and V4, the middle temporal area (MT), and MST as well as the caudal part of STS (Blatt et al., 1990); it has also been shown to send projections to the intermediate and deep layers of the SC (Andersen, 1997). This area also sends dense projections to frontal eye fields (Blatt et al., 1990), and electrical stimulation of LIP evokes saccadic eye movements (Mountcastle et al., 1975). Unlike those of the adjacent visual area 7a, the visual receptive fields of neurons in LIP are small and almost always located in the contralateral hemifield (Blatt et al., 1990). In the context of a delayed-saccade task, visually-responsive neurons in LIP are active in a direction-selective manner during stimulus presentation as well as during the delay preceding response (Andersen et al., 1990). Importantly, a subset of neurons in this area is also active during an auditory version of this task, and most auditory-responsive cells appear to code for the location of the auditory stimulus in a manner similar to that seen in the SC (Mazzoni et al., 1996; Stricanne et al., 1996). In combination with evidence of eye- and head-centered gain fields in LIP neurons (Bradley et al., 1996), and recent work demonstrating an ability to determine stimulus location based upon single-neuron activity in LIP (Mullette-Gillman et al., 2005), this work points to a role for this region in transforming a

multimodal representation of space into a motor coordinate framework appropriate for orientation and action.

The macaque ventral intraparietal area (area VIP) lies in the fundus of the intraparietal sulcus, adjacent to area LIP. In addition to receiving abundant projections from area MT (Maunsell and van Essen, 1983), VIP has been shown to receive significant projections from a wide variety of visual, somatosensory, motor, auditory, vestibular, and multisensory cortical areas (Lewis and Van Essen, 2000). This is in striking contrast to LIP, which predominantly receives projections from visual area MT (Lewis and Van Essen, 2000). In accordance with this pattern of anatomical connectivity, neurons in VIP are responsive to visual, auditory, somatosensory, and vestibular stimulation—with estimates of ~70% of neurons being at least bimodal (Duhamel et al., 1998)—and the constitutive receptive fields of bi- and tri-modal neurons in this area exhibit a great degree of overlap and similar directional selectivity (Colby et al., 1993; Schlack et al., 2002; Schlack et al., 2005). In a striking similarity to the integrative properties of SC neurons, VIP neurons have recently been shown to exhibit response enhancements and depressions as well as a shortening of response latency when presented with congruent visuo-tactile stimulus combinations (Avillac et al., 2007). Intriguingly, subsets of neurons in VIP seem to encode not only visual or vestibular motion in general, but self-motion in particular (Gabel et al., 2002), leading some to propose that the area is important in spatial navigation. While the function of multisensory integration in VIP is still being debated, the multimodal representation of peripersonal space (Ladavas and Farne, 2004) via coordinate transformation of inputs from sensory and motor systems (Bremmer et al., 1999) into a common framework, the

guidance of head movements (Duhamel et al., 1998), and navigation of space in general (Bremmer, 2005) remain notable possibilities.

Studies of human posterior parietal cortex reinforce these views of the region's importance in constructing multimodal map of extrapersonal space and in navigating that map successfully. For decades, lesions in right posterior parietal cortex have been known to produce not only contralateral hemispatial neglect, but also somatoparaphrenia, or a disowning of the body parts of half of one's body (Vallar and Ronchi, 2009). Underscoring its importance for exploration of peripersonal space, several recent studies have indicated that posterior parietal cortex is essential for haptic exploration of objects and visuo-haptic matching (Deshpande et al., 2008; Miquee et al., 2008; Lacey et al., 2009). The area also appears to be involved in the visual override of proprioceptive input when evaluating limb motion (Hagura et al., 2007), and also in the integration of visual and proprioceptive inputs during reaching behaviors within so-called parietal reach regions (PRR) (Andersen et al., 1998; Snyder et al., 2000; Filimon et al., 2009), as well as in evaluation of auditory-visual spatial congruity (Lewald et al., 2002; Meienbrock et al., 2007). Like STS, posterior parietal cortex has also been implicated in multisensory temporal processing. Non-human primate work in LIP has shown that neurons in this region respond preferentially during active comparisons of temporal duration (Leon and Shadlen, 2003), and the area appears to be involved specifically in the detection of audiovisual synchrony/asynchrony (Bushara et al., 2001; Dhamala et al., 2007). Overall, however, the human literature on multisensory integration in posterior parietal regions recapitulates what is known from work in non-human primates: this region is essential

for the performance of coordinate transformations across sensory and motor systems and for the accurate assessment and navigation of immediate extrapersonal space.

Frontal and Prefrontal Regions

Several frontal and prefrontal regions have also been shown to be important multisensory convergence areas. By comparison to other putative loci of multisensory integration, however, relatively few studies have focused on these regions. Monkeys trained in both auditory frequency discrimination and color discrimination have been shown to exhibit prefrontal neural activity in response to both visual and auditory stimuli (Fuster et al., 2000; Deco et al., 2005). Moreover, neurons in ventrolateral prefrontal cortex (VLPFC) have been shown to integrate visual and auditory vocal signals, and their integration is known to be contingent upon the congruity of the information presented (Sugihara et al., 2006; Romanski, 2007). Successful lipreading has been shown to activate this region in humans as well (Kang et al., 2006). From this evidence alone, VLPFC seems to be emerging as a locus of high-level audiovisual stimulus matching. More posterior frontal regions, on the other hand, appear to be involved in detection of and defense against approaching objects. In particular, premotor region F4 responds preferentially to looming visual, auditory, and somatosensory stimuli in the vicinity of the upper torso, face, and arms (Graziano et al., 1994; Fogassi et al., 1996; Graziano et al., 1999). Consistent with the interpretation that this region is important for defense against approaching stimuli, microstimulation of this area has been shown to induce defensive-like movements (Graziano et al., 2002). Very recent human studies have further characterized the multisensory features of this region: viewing one's own face being

touched decreased BOLD activity in premotor regions and somatosensory cortex in comparison with a touch-alone condition (Cardini et al., 2010); premotor cortex in the region of Broca's area has been identified as being important for comprehension of audiovisual speech in noise (Bishop and Miller, 2009); finally, activity in premotor cortex has been shown to correlate with perception of the so-called rubber hand illusion, wherein simultaneously stroking a rubber hand and one's own hand results in a feeling of ownership of the rubber hand (Ehrsson et al., 2004). Overall, multisensory integration in frontal and prefrontal regions appears to reflect higher-order synthesis of information from different sensory systems.

Multisensory Integration in "Unisensory" Cortex

While the above analysis of multisensory integration in higher-order cortical regions fits well with the traditional hierarchical view of sensory processing, an emerging body of evidence indicates that multisensory interactions may occur much earlier, in regions previously thought to be the exclusive domain of unisensory stimuli. First, it has become clear that activation of auditory cortex with visual or somatosensory stimuli alone is possible: early work using fMRI to study the effect of silent lipreading on auditory areas showed that a silent video of human speech (as contrasted against non-speech facial gestures) is capable of activating primary auditory cortex (Calvert et al., 1997). Similarly, work analyzing event-related potentials (ERPs) in response to audiovisual speech indicates that the addition of visual input to a congruent auditory speech stream is capable of speeding the initial auditory processing of that stream (van Wassenhove et al., 2005), suggesting that activity in auditory cortex can be influenced by

visual input even at the earliest stages of auditory processing. During simultaneous recording of single-unit, multi-unit, and local field potentials in macaque auditory cortex, Schroeder and colleagues demonstrated that somatosensory-driven local field potentials in auditory cortex have similar latencies to those of auditory-driven potentials and exhibit a feed-forward laminar profile (Schroeder and Foxe, 2002). More recently, a high-resolution fMRI study has demonstrated that concurrent broadband noise and tactile stimulation is capable of producing enhanced BOLD signals in caudal auditory belt when compared to auditory presentation alone, and that this pattern of enhancement obeys the principle of inverse effectiveness (Kayser et al., 2005).

The effect of non-visual stimuli on visual cortex has similarly been demonstrated. Early work by Morrell (Morrell, 1972) recorded single-unit activity in cat BA 18 and 19 and found that cells were responsive to both auditory (pure tones, click trains, and noise bursts) and visual stimuli, and that visual and auditory receptive fields were co-registered. Later work has not only replicated these findings, but has demonstrated that non-visual inputs are capable of affecting very early visual processing. In the context of an audiovisual reaction-time task, a study using high-density electrical mapping has demonstrated very early (~46 ms) audiovisual interactions in lateral occipital cortex, coincident with the earliest signature of visual cortical processing in that region (Molholm et al., 2002). Moreover, shifts in visual temporal order judgment brought about by exogenous attentional shifts to an ipsilateral auditory cue have been shown to be correlated with increases in amplitude of early visual ERP waveforms (McDonald et al., 2005). Somatosensory influences on visual cortex have similarly been demonstrated: haptic object identification activates lateral occipital cortex (Amedi et al., 2001; James et

al., 2002), and perception of tactile motion is known to activate human area MT (Hagen et al., 2002; Blake et al., 2004; Ptito et al., 2009).

In contrast to the amount of evidence of cross-modal influences on auditory and visual cortices, there appear to be relatively few studies that have investigated non-somatosensory influences upon early somatosensory cortex. Electrophysiological studies in monkeys trained to make visuo-haptic or audio-haptic associations have shown that a subset of cells in primary somatosensory cortex respond both to the somatosensory and associated visual or auditory cues (Zhou and Fuster, 2000, 2004). Recent studies by Schaefer and colleagues indicate that perception of the rubber hand illusion may be accompanied by shifts in the location of the S1 hand representation (Schaefer et al., 2006; Schaefer et al., 2009), but the functional significance of these shifts has yet to be determined.

The anatomical and physiological underpinnings of these early multisensory interactions are the subject of continuing debate. There is ample evidence of feedback projections from multimodal cortical regions to traditional unisensory regions as well as lateral connections among unisensory regions. Primary visual cortex has been shown to receive sparse inputs from A1 and heavy projections from auditory belt regions as well as from STS (Falchier et al., 2002; Rockland and Ojima, 2003), most of which project to areas responsive to the peripheral visual field (Falchier et al., 2002) and terminate on cortical layers 1 and 6, in a classic feedback-style pattern (Rockland and Ojima, 2003). Evidence of feedback projections to auditory areas has been less plentiful, but a study has recently demonstrated visual projections arising from an area anterior to STS and terminating in core regions of auditory cortex (Cappe and Barone, 2005). Furthermore, if

the projections described by Falchier and colleagues follow the typical cortical pattern of reciprocal projection (Rockland and Pandya, 1979), more instances of feedback from STS to auditory regions may soon be described. Finally, while examples of auditory or visual activation of somatosensory cortex are sparse, anatomical studies in marmosets have revealed projections from visual areas MT and FST to somatosensory regions I and 3b as well as from auditory regions to secondary somatosensory cortex (Cappe and Barone, 2005).

In addition to these feedback pathways, feed-forward pathways have been increasingly described in the literature as a means by which early multisensory integration may occur. The case for feed-forward connectivity has in large part been made based upon non-primary single-unit response latencies in visual, auditory, and somatosensory areas that are too short to reflect feedback input (Schroeder and Foxe, 2002; Foxe and Schroeder, 2005; Lakatos et al., 2005; Chen et al., 2007; Lakatos et al., 2007; Musacchia and Schroeder, 2009). The possible sources of this feed-forward input are continuing to be debated, but several anatomical substrates have been proposed. Subcortical nuclei such as the thalamic posterior (PO), ventral posterior (VP), postero-medial (PO), limitans (LIM), and superageniculate (SG) nuclei, in addition to the magnocellular division of the medial geniculate nucleus, have been cited as possible sources for projection of somatosensory and visual inputs to auditory cortex (de la Mothe et al., 2006b, a; Hackett et al., 2007), and the importance of pulvinar as a possible mediator of feed-forward processing is beginning to be recognized (Sherman and Guillery, 2002; Cappe et al., 2009a). Lastly, the role of even earlier subcortical structures in these circuits cannot be ruled out: auditory-somatosensory interactions have

been shown to occur as early as the dorsal cochlear nuclei (Young et al., 1995; Davis et al., 1996; Shore, 2005).

From this evidence, it is clear that cortical multisensory convergence areas, unisensory cortices, subcortical sensory structures, and their interconnectivity play essential roles in multisensory processing, and that a network-based approach to understanding how multisensory interactions occur will be essential in moving forward. New methods of investigation meant to assess the dynamics of distributed networks are now being developed, building off of tools traditionally used to assay the activity of individual brain areas (McIntosh et al., 1994; Horwitz et al., 1995; Horwitz et al., 1999). For example, analysis of EEG and local field potential data has provided a robust tool for characterization of functional connectivity among different brain regions (Koenig et al., 2005). In this methodology, neuronal synchrony detected by different frequency oscillations on EEG has revealed dynamic interactions among brain regions thought to underlie various perceptual and cognitive functions (Fries, 2005; Hong et al., 2005; Bastiaansen and Hagoort, 2006; Senkowski et al., 2007; Masuda, 2009). A similar approach has been taken with the use of fMRI, detecting small deflections in BOLD signal thought to reflect activity of ensembles of neurons during rest and determining if this activity is correlated across voxels. This particular approach to assessing functional connectivity has evolved significantly of late: in addition to traditional approaches assessing such BOLD correlations on a whole-brain level (Friston et al., 1993; Friston, 1994), newer methods have arisen that take anatomy into account in constraining hypotheses about functional connectivity. For example, structural equation modeling (SEM) uses Granger causality analysis of BOLD time series to make inferences about

effective connectivity among nodes (MacCallum and Austin, 2000). Similarly, dynamic causal modeling (DCM) allows investigators to specify the nodes within a network, driving influences upon them within the context of a task, and the way they might interact in a series of models (Lee et al., 2006). Evidence for those models is then assessed given functional (i.e., EEG, MEG, fMRI) data and the posterior probability that a model accounts for the given data is then calculated via Bayesian model selection (Stephan et al., 2007; Stephan et al., 2010). Results produced by such an analysis are included in Chapter III, and a discussion on the use of functional and effective connectivity techniques in current and future investigations forms a major part of Chapter IV.

Regardless of its anatomical and neurophysiological substrates, however, multisensory integration has been shown to conform to certain principles in determining how inputs are combined to produce an array of physiological, perceptual, and behavioral enhancements. These rules are outlined below.

Principles of Multisensory Integration

Multisensory integration, as first examined in the SC, seemed to produce the largest enhancements when certain criteria were met (Meredith and Stein, 1983; Meredith et al., 1987; Stein and Meredith, 1993). These criteria, later termed principles of multisensory integration, are concerned with the spatial and temporal congruence of multisensory stimuli as well as their effectiveness in eliciting responses when presented alone. Here we briefly review their derivation and possible behavioral relevance before focusing directly upon the temporal principle as a main point of interest for this thesis.

Spatial Congruity

Multisensory neurons in the deep layers of the SC have several receptive fields, corresponding to each of the modalities to which they respond (Stein and Arigbede, 1972; Meredith and Stein, 1983). These receptive fields are spatially co-registered (e.g., the auditory and visual receptive fields of an audiovisual-responsive neuron overlap in space). Indeed, if the information coming from each modality is presented within the space representing their respective and overlapping receptive fields, combined presentation will often lead to response enhancement and if one stimulus is presented within and one outside of their respective receptive fields, their combination often leads to response depression (Meredith and Stein, 1983; Kadunce et al., 2001). This pattern of responses may well be due to the architecture of the unimodal receptive fields themselves: a flash-tone pair, for example, if presented within the individual auditory and visual receptive fields, will produce response enhancement. However, if one stimulus is displaced so that it is presented within the inhibitory region surrounding its own receptive field while its counterpart is still within its own excitatory receptive field, the combined excitatory and inhibitory influences may result in an overall response depression (Kadunce et al., 2001). While this model is conceptually appealing, the heterogeneity inherent in multisensory receptive fields hints at a far more complex relationship (Carriere et al., 2008; Krueger et al., 2009; Royal et al., 2009; Royal et al., 2010).

The importance of spatial congruence on multisensory integration has also been demonstrated in human psychophysical studies. The shortening of saccade latencies

produced when visual targets are paired with auditory stimuli is contingent upon spatial congruity (Frens et al., 1995), and the influence of congruent visual input upon discrimination of auditory motion has been well-documented (Soto-Faraco et al., 2004). Moreover, manipulation of spatial congruity has been shown to alter the performance of audiovisual simultaneity and temporal order judgment tasks (Zampini et al., 2003; Keetels and Vroomen, 2005; Zampini et al., 2005b), with simultaneity typically being reported at higher rates and across more SOAs with spatially-aligned stimuli.

Manipulation of multisensory spatial congruity has proven very useful in the study of cross-modal bias. Perhaps best known of these examples is the case of the ventriloquist effect, wherein localization of an auditory cue is biased by concurrent presentation of a visual cue (Howard and Templeton, 1966). Studies of this effect have demonstrated that bias is typically greatest at central locations in the visual field and with relatively small degrees of auditory-visual spatial disparity (Bertelson et al., 2000; Lewald et al., 2001; Slutsky and Recanzone, 2001; Hairston et al., 2003). Interestingly, the decrease in bias observed with large disparities is often accompanied by decreases in perception of spatial unity of the stimuli, and the degree of bias observed was seen to covary with perception of unity, supporting the notion that the effect itself results from the constituent cues being perceptually bound into a common percept (Bertelson and Radeau, 1981). It should be noted that these influences are not particular to audiovisual interactions, however. In a surprising study of the effect of proprioceptive input upon localization of visual stimuli, stimulation of neck muscles was capable of biasing apparent visual stimulus location (Taylor and McCloskey, 1991). From these examples,

it is clear that spatial congruity is an important factor in the integration of cross-modal stimuli.

Stimulus Effectiveness

The influence of stimulus effectiveness on the degree of multisensory enhancement observed is another key organizing principle of multisensory integration. As such, the magnitude of enhancement seen is inversely proportional to the effectiveness of the individual stimuli when presented alone (Meredith and Stein, 1983; Meredith and Stein, 1986a; Wallace et al., 1996). This relationship, having been first uncovered in the SC, has clear implications for stimulus detection and orientation behaviors: less salient individual environmental events are far more likely to generate an orientation behavior in combination. Interestingly, this principle seems to apply within multisensory spatial receptive fields themselves, predictably driving enhancements and depressions based upon stimulus placement within heterogeneous unisensory spatial receptive fields (Carriere et al., 2008; Krueger et al., 2009). As a principle, inverse effectiveness has been shown to apply to human psychophysical studies of understanding audiovisual speech in noise (Sumbly and Pollack, 1954; Ma et al., 2009), multisensory-mediated auditory stimulus localization (Bolognini et al., 2005; Bolognini et al., 2007), as well as BOLD measures of multisensory integration (Kayser et al., 2005; Cappe et al., 2009b; Stevenson and James, 2009; Stevenson et al., 2009; Werner and Noppeney, 2009; Holle et al., 2010).

Like the dependence of multisensory integration upon the spatial congruity of stimuli, temporal congruity also governs the degree to which cross-modal stimuli are

perceptually bound. Because this principle is central to the present work, it is considered separately in the following section.

*The Temporal Principle,
The Multisensory Temporal Binding Window,
And Cross-modal Simultaneity Perception*

On the level of the single cell, early studies made clear that multisensory integration in SC neurons is dependent upon stimulus onset asynchrony (SOA). While individual neurons in cat SC and cortex appear to be tuned to different SOAs, the span in time over which response enhancements are generally seen in these neurons is on the order of several hundred milliseconds and depends upon the degree of overlap of the individual unisensory response trains (Meredith et al., 1987; Wallace et al., 1992; Wallace and Stein, 1996; Wallace et al., 1996). Beyond this elementary work, however, the so-called temporal principle has been studied most commonly in behavioral work. Presentation of cross-modal stimulus pairs in close temporal proximity has been shown to be important for shortened saccadic reaction times (Frens et al., 1995; Colonius and Arndt, 2001; Colonius and Diederich, 2004), heightened accuracy in understanding speech in noise (McGrath and Summerfield, 1985; Pandey et al., 1986; van Wassenhove et al., 2007), as well as mediating multisensory illusions such as the McGurk effect (Munhall et al., 1996), the sound-induced flash illusion (Shams et al., 2000, 2002), the parchment skin illusion (Guest et al., 2002), and the stream-bounce illusion (Sekuler et al., 1997). Moreover, multisensory interactions as demonstrated using functional imaging (Macaluso et al., 2004; Dhamala et al., 2007; Noesselt et al., 2007;

Kavounoudias et al., 2008) and ERPs (Schroeder and Foxe, 2002; Tanaka et al., 2009) have been shown to be greatest during synchronous presentation of stimulus pairs. However, as was seen in single-unit electrophysiology, multisensory interactions in behavior and functional imaging do not depend upon absolute synchrony, but are demonstrable over a range of SOAs spanning several hundred milliseconds. This observation led several researchers to describe the temporal dependence of multisensory integration in terms of a multisensory temporal binding window within which cross-modal stimulus pairs are capable of producing behavioral, perceptual, and electrophysiological benefits (Dixon and Spitz, 1980; Colonius and Diederich, 2004; van Wassenhove et al., 2007).

Because temporal correspondence appears to be so critical for multisensory integration, several experimental paradigms have been developed for the systematic study of cross-modal simultaneity perception as a proxy for the temporal binding window. In a simultaneity judgment task, pairs of visual and auditory stimuli are presented at a range of SOAs and participants are asked to judge whether the stimuli occurred simultaneously or successively (Engel and Dougherty, 1971; Stone et al., 2001; Zampini et al., 2005b; Stevenson et al., 2010). Responses are then plotted as a function of SOA and the point of subjective simultaneity (PSS) is derived as the peak of function (Stone et al., 2001; Zampini et al., 2005b). A similar approach is taken in the use of cross-modal temporal order judgments, wherein participants judge whether stimuli within one or another modality was presented first, and the PSS is the time point at which participants judge either stimulus to have occurred first at a rate of fifty percent (Spence et al., 2001; Zampini et al., 2003). Other studies have used detection of synchrony/asynchrony in an

attempt to identify the range of SOAs over which auditory and visual streams are perceived as synchronous (Dixon and Spitz, 1980; McGrath and Summerfield, 1985). While the point measures used here tend to differ based upon the paradigm chosen (Zampini et al., 2003; Fujisaki et al., 2004; Vroomen et al., 2004; Zampini et al., 2005b), the span in time over which the likelihood of reporting simultaneity is remarkably constant, ranging from about -100 ms to 250 ms, where negative values denote auditory-leading-visual conditions (Dixon and Spitz, 1980; Zampini et al., 2003; Fujisaki et al., 2004; Vroomen et al., 2004; Zampini et al., 2005b). The larger window size on the right side of these distributions—in which vision leads audition—appears in nearly all studies of audiovisual simultaneity perception, and has been proposed to arise from the inherent flexibility needed to process real-world audiovisual events, given that the propagation speeds of light and sound will result in SOAs only on the right side of these distributions (Dixon and Spitz, 1980).

Despite the relative consistency of these measures across studies, several factors have been shown to affect the temporal binding window. First, attending to one modality has been shown to speed perception of stimuli in that modality, a phenomenon termed *prior entry* (Stevens, 1904; Shore et al., 2001; Spence et al., 2001; Zampini et al., 2005a). Experimental evidence of this phenomenon typically takes the form of PSS shifts on cross-modal temporal order judgment tasks wherein attention is manipulated (Spence et al., 2001). Over the years, results of experiments designed to demonstrate prior entry have been ambiguous. Whereas initial reports found little evidence (Hamlin, 1895; Drew, 1896), later studies generated ample data but were confounded by the effects of spatial attention (Frey, 1990; Jaskowski, 1993). Recent studies in a multisensory framework

have indeed confirmed that attention to one modality can bias temporal order judgments in the direction of the attended stimulus (Shore et al., 2001; Spence et al., 2001; Zampini et al., 2005a), but whether these results address the ability of attention to speed the perception of an event is still debatable. In fact, some work utilizing ERPs (which have excellent temporal resolution) has found very little evidence that prior entry effects can be explained by speeding of early sensory responses (Schneider and Bavelier, 2003), while others have found evidence of a modest shift in peak P1 latencies when visual events are attended (Vibell et al., 2007). Regardless, there is clearly a role for attention in mediating responses during cross-modal temporal order judgment tasks. This evidence, together with the data described above supporting cross-modal effects in unisensory cortices, strongly supports the idea that even early sensory processing is susceptible to top-down influences.

The type of stimulus used has also been shown to have a profound effect on participants' likelihood of detecting asynchrony at any given SOA, with speech stimuli typically producing a far larger temporal window—about 450 ms—than flash-tone pairs or videos of objects performing some action, such as a hammer pounding a nail—about 250 ms (Dixon and Spitz, 1980; McGrath and Summerfield, 1985; Massaro et al., 1996; Conrey and Pisoni, 2006; van Atteveldt et al., 2007; van Wassenhove et al., 2007). Interpretation of this seeming expansion in the case of speech has ranged from the idea that learned tolerance of asynchrony is greatest with stimuli to which we are most exposed (Dixon and Spitz, 1980) to the theory that the richness of auditory spectral and visual dynamic content in speech allows for binding over a larger range of asynchrony (Massaro et al., 1996). Along these lines, studies demonstrating that the same tolerance

to asynchrony is observed when lip movements are mimicked with the use of Lissajous figures composed of dot arrays indicate that this tolerance may rely heavily upon dynamic lip movements (Massaro and Cohen, 1993). Later studies decomposing auditory speech signal into low- (298-375 Hz) and high-frequency (4762-6000 Hz) bands further indicate that even these reduced signals are capable of producing an increased tolerance to asynchrony relative to studies using only simple stimuli (Grant and Greenberg, 2001).

Because the ethological arguments for the temporal binding window's asymmetry have focused upon the different propagation and arrival times of light and sound at a distance, several studies have sought to determine if the brain takes distance into account when judging audiovisual simultaneity, thus producing simultaneity constancy regardless of distance (Engel and Dougherty, 1971). Studies on this subject have indicated that participants do take the distance of audiovisual events into account when judging their simultaneity; thus, participants' PSS values shift toward increasingly asynchronous values, partially—or by some accounts, wholly—compensating for the difference in arrival times of auditory and visual stimuli and leading to perception of simultaneity at the distance at which the stimuli were produced (Engel and Dougherty, 1971; Kopinska and Harris, 2004). Once again, these results support the idea that multisensory processing may be influenced by top-down factors.

The flexibility of the PSS point measure has also been demonstrated in its sensitivity to recalibration after repeated exposure to asynchronous stimuli. Specifically, several studies have now shown that participants are far more likely to judge asynchronous cross-modal pairs to be simultaneous after repeated exposure to asynchronous pairs. Thus, repeated exposure to an 250-ms auditory-leading-visual

asynchronous pair is capable of biasing participants' PSS in the direction of that lag by about 25 ms, with effects lasting on the order of minutes (Fujisaki et al., 2004; Vroomen et al., 2004). Similar recalibration effects have been noted after exposure to asynchronous audiovisual speech, as well as to visual-tactile, audio-tactile, and sensory-motor pairs (Navarra et al., 2005; Stetson et al., 2006; Fajen, 2007; Hanson et al., 2008). While the exact mechanisms underlying this phenomenon are unknown, they have been proposed to represent a recalibration of sensory input that does not match prior expectations, consistent with Bayesian models of perception (Miyazaki et al., 2005; Miyazaki et al., 2006; Hanson et al., 2008). This hypothesis has recently been supported—albeit indirectly—by data from functional imaging demonstrating increased BOLD signal after sensory-motor recalibration in anterior cingulate cortex, which has been shown to be critically involved in conflict monitoring (Stetson et al., 2006).

This work has shown that perception of audiovisual simultaneity may be updated dynamically, but the capability of the multisensory temporal binding window and adult multisensory systems in general to exhibit plastic change has not yet been investigated. This will be the focus of the remainder of this volume, representing the first demonstration of adult multisensory systems' capability for lasting plastic change. In formulating the work described in the following chapters, we were able to draw from an extensive literature describing sensory plasticity both during the course of development and in the adult. While a comprehensive review of this topic is beyond the scope of this introduction, the following section provides a brief review of major themes in sensory plasticity.

Themes in Sensory Plasticity

Adaptive sensory plasticity is essential for tuning neural circuits to properties of the world that must be learned from experience. In the realm of development, this plasticity is unusually strong during early sensitive periods, wherein sensory experience is capable of permanently altering neuronal response properties and organization. After the closure of these sensitive periods, plasticity is most often observed as compensatory changes after neural injury and adaptive plasticity wherein changes on the neural level reflect a change in the behavioral relevance of sensory stimuli. In this section, tenets of developmental and adult unisensory plasticity are first reviewed, followed by a brief review of adaptive plasticity in the multisensory realm.

Plasticity in Developing Animals

Classical studies of developmental plasticity grossly manipulated sensory experience during sensitive periods of development and observed the ensuing organizational changes in sensory cortex. In the visual system, this was most extensively seen as reorganization of ocular dominance columns after monocular deprivation, which has been described in cat, rat, ferret, monkey, and human (Hubel and Wiesel, 1963, 1968; Banks et al., 1975; Hubel et al., 1977; Fagiolini et al., 1994; Issa et al., 1999; Bengoetxea et al., 2008). Similar work in the auditory and somatosensory systems has focused upon monaural plugging, cochlear lesions, and digit/whisker removal or denervation (Kaas et al., 1983; Merzenich et al., 1983; Diamond et al., 1993; Samson et al., 1993; Popescu and Polley, 2010). Collectively, these studies have resulted in the basic tenet that the organization (and reorganization) of sensory neural circuits is driven by competing

inputs. Thus, when input from one eye is removed by enucleation or monocular deprivation, its share of cortex becomes responsive to input from the other, active eye. In contrast, in the case of binocular deprivation, the cortex remains in an immature state (i.e., the sensitive period is effectively extended) and near-normal development of ocular dominance columns is seen after normal sensory input is restored (Hubel and Wiesel, 1963; Wiesel and Hubel, 1965; Hubel et al., 1977). Consistent with the concept of competition, deprivation and over-exposure produce very similar results: selective presentation of a 4-kHz tone during development of a rat has been shown to increase the representation of that frequency in primary auditory cortex (Zhang et al., 2001), and early over-stimulation of one digit relative to another produces a similar increase of its representation in somatosensory cortex (Simons and Land, 1987). Effects of early exposure and the existence of sensitive periods for more complex sensorimotor and perceptual phenomena, such as bird song, human language learning, and musical training have also been demonstrated (Johnson and Newport, 1989; Pantev et al., 1998; Doupe and Kuhl, 1999). During the course of development, then, it is clear that manipulation of the sensory environment is capable of eliciting profound and lasting changes in the neural structure of sensory systems.

Plasticity in Adults

Despite the importance of sensitive periods for the developmental plasticity observed in sensory systems, it has become increasingly clear that adult systems are also capable of plastic change. The earliest studies of adult plasticity actually focused upon inter-sensory compensatory changes, wherein the loss of input from one modality

seemingly increased sensitivity and/or neural responsiveness to another (Kellogg, 1962; Rice et al., 1965; Korte and Rauschecker, 1993; Rauschecker and Korte, 1993). Later, cortical map reorganization after digital deafferentation was shown to occur in adult animals well past their developmental critical periods (Merzenich et al., 1983).

Intact adult sensory systems must be able to adapt to constantly changing environmental demands as well, but they must be stable enough to allow for reliable sensory processing, achieving what some have termed the *plasticity-stability balance* (Grossberg, 1980; Ogasawara et al., 2008). As such, sensory systems must be able to determine when plastic change is necessary and which features of new environments should drive change. Such a determination is central to perceptual learning, in which perceptual abilities improve with practice. Most commonly, perceptual learning is studied in one of two ways. One might look for differences in perception between a group with particular learned skills (e.g., birdwatchers, radiologists, wine experts, etc.) and those who have not had the training necessary for development of those skills (Bende and Nordin, 1997; Gauthier et al., 2000; Gauthier et al., 2003). In a second approach, training on a perceptual task may be undertaken in a laboratory setting and changes in perception and/or neural function assessed after that training. Such an approach forms the basis of the studies described in this volume, and has been used to demonstrate perceptual improvements in visual (Fiorentini and Berardi, 1980; Adini et al., 2002), auditory (Annett, 1966; Wright et al., 1997; Polley et al., 2004), somatosensory (Spengler et al., 1997; Nagarajan et al., 1998), and olfactory (Stevenson, 2001; Wilson and Stevenson, 2003) discrimination abilities. In a visual example, one's ability to discriminate between lines of different orientations improves robustly with training over

the course of several sessions (Schoups et al., 2001). These improvements are typically stable over time, lasting months (Watanabe et al., 2002) or sometimes years (Karni and Sagi, 1993), and are often specific to particular stimulus parameters—such as location in space or contrast (Adini et al., 2002)—although some notable exceptions do exist (Liu, 1999; Mossbridge et al., 2008; Xiao et al., 2008; Zhang et al., 2010).

The neural changes underlying perceptual learning have been the subject of some debate. While it makes intuitive sense that such improvements would be driven by changes in sensory areas, recent evidence indicates that they could also result from improvements in the way that sensory representations are interpreted in decision-making areas (Law and Gold, 2008). Most work has supported the former hypothesis, however: perceptual learning-driven plasticity has been demonstrated in visual, somatosensory, and auditory cortices as well as in subcortical sensory structures (Fiorentini and Berardi, 1980; Diamond et al., 1999; Pleger et al., 2001; Skrandies et al., 2001; Foltz, 2003; de Boer and Thornton, 2008). Robust and predictable cortical reorganization has been demonstrated in cases of perceptual learning, wherein a particular stimulus or stimulus property is consistently paired with reward or punishment, thus gaining increasing behavioral salience (Linkenhoker and Knudsen, 2002; Bergan et al., 2005; Polley et al., 2006). Several studies have demonstrated that focused attention and conscious effort are required to drive certain types of perceptual plasticity (Shiu and Pashler, 1992; Ahissar and Hochstein, 1993). The importance of the top-down effects of attention, effort, and reward forms the basis of the hypothesis that behavioral relevance, signaled by changes in neuromodulatory signalling, allows mature circuits to overcome their inherent stability and produce plastic change. This idea has been supported by a multitude of

recent studies demonstrating that acetylcholine agonists and antagonists are capable of modulating or preventing the occurrence of perceptual learning (Ji et al., 2001; Ji and Suga, 2003; Weinberger, 2004; Ji and Suga, 2008) and that dopaminergic or cholinergic modulation via electrical stimulation of the ventral tegmental area or nucleus basalis is capable of regulating auditory plasticity (Bakin and Weinberger, 1996; Kilgard and Merzenich, 1998b, a; Bao et al., 2001; Kilgard and Merzenich, 2002; Bao et al., 2003; Kilgard, 2003). The exact nature of the cellular and/or network changes wrought by these influences on sensory cortex is not known at this time, but it is clear that both signal enhancement by Hebbian synaptic strengthening (Poggio et al., 1992; Fahle et al., 1995; Fahle, 2004) and noise reduction by lateral inhibition (Crist et al., 2001; Schwartz et al., 2002; Hoshino, 2004) play key roles.

Multisensory Plasticity

Despite the richness of data on plasticity in all of the sensory systems, the study of multisensory plasticity is in its nascent stages. Like the study of unisensory developmental plasticity, multisensory plasticity in development has been studied most extensively in the context of sensory deprivation, alteration, or injury and its effects on multisensory function (Calvert et al., 2004; Wallace and Bear, 2004; Carriere et al., 2007; Polley et al., 2008; Roder and Wallace, 2010). Early neurophysiological studies demonstrated compensatory auditory changes after visual deprivation in the deep—and even superficial—layers of the SC (Rauschecker and Harris, 1983), and similar compensatory plasticity was observed in visual cortical area 19 (Hyvarinen et al., 1981). Shorter latencies for auditory and somatosensory ERP waveforms have been shown in

blind when compared with sighted individuals (Feinsod et al., 1973; Niemeyer and Starlinger, 1981). Consistent with the idea that occipital cortex can be co-opted for processing of other sensory information in the blind, visual cortical activity has been demonstrated in the blind during somatosensory and auditory tasks (Neville et al., 1983; Sadato et al., 1996; Kujala et al., 1997; Bavelier et al., 2001; Burton et al., 2006), and the converse has been shown in auditory cortex of congenitally deaf individuals during visual and somatosensory tasks (Finney et al., 2001; Bavelier and Neville, 2002; Fine et al., 2005). Strikingly, this reorganization reveals a true shift in function: studies using transcranial magnetic stimulation (TMS) to create functional lesions in early-blind participants during auditory spatial localization tasks have demonstrated that TMS-mediated disruption of occipital cortex (but not auditory cortex) is capable of degrading performance (Collignon et al., 2009b). The search for increased acuity in one sense after deprivation in another has yielded more ambiguous results, however: no differences in auditory frequency discrimination have been detected between blind and sighted individuals (Niemeyer and Starlinger, 1981; Starlinger and Niemeyer, 1981; Bross and Borenstein, 1982), but auditory gap detection, localization, and general temporal processing have been shown to be improved in the blind (Muchnik et al., 1991; Rammsayer and Vogel, 1992; Collignon et al., 2009a). Similarly, while overall haptic sensitivity has been shown to be unaltered in the blind (Pascual-Leone and Torres, 1993), tactile hyperacuity (Grant et al., 2000) and grating orientation tasks (Van Boven et al., 2000) yield superior performance in blind Braille readers.

Moving beyond the realm of compensatory changes, several recent developmental studies have shown that while multisensory neurons in the SC (Wallace et al., 2004) and

in the anterior ectosylvian sulcus of dark-reared cats (Carriere et al., 2007) retain their visual responsiveness, their unique integrative properties are lost. Even more strikingly, alteration of multisensory spatial co-registration during development is capable of altering the spatial alignment of receptive fields of SC neurons while retaining their ability to integrate that information (Wallace and Stein, 2007), indicating that exposure to altered environmental statistics during development is capable of eliciting lasting change in multisensory systems.

Multisensory processes are also capable of change into adulthood, although most studies have not examined multisensory perceptual learning *per se*, but have focused instead upon facilitation of unisensory perceptual learning with the use of cross-modal stimuli. For example, several studies have trained different groups of participants on audiovisual and visual-alone versions of a motion discrimination task. Results reveal enhanced visual motion discrimination abilities and an abbreviated time course of learning in the group trained on the audiovisual version of the task when compared with those trained only on the visual version (Seitz et al., 2006; Kim et al., 2008). Similar results have been seen in the visual facilitation of voice discrimination learning (von Kriegstein and Giraud, 2006) and cross-modal facilitation of both auditory and visual natural object recognition (Schneider et al., 2008).

Introduction to the Current Work

The examples above illustrate that multisensory processes are capable of being shaped by the sensory environment during development, and that unisensory perceptual learning is susceptible to cross-modal influences in the adult. However, no study to date has examined the ability of adult multisensory systems themselves to be altered with perceptual learning. The work contained in Chapter II represents the first attempt to characterize the multisensory temporal binding window with the use of an audiovisual simultaneity judgment paradigm and to narrow it with the use of perceptual training. This is followed directly by Chapter III, which offers evidence that this narrowing is the result of functional changes in multisensory posterior superior temporal sulcus as well as in auditory and visual areas. These changes are described in conjunction with alterations in functional and effective connectivity among these areas after training. Lastly, the General Discussion comprising Chapter IV summarizes the major findings of this work, discusses pertinent themes, and proposes directions for future study.

References

- Adini Y, Sagi D, Tsodyks M (2002) Context-enabled learning in the human visual system. *Nature* 415:790-793.
- Ahissar M, Hochstein S (1993) Attentional control of early perceptual learning. *Proc Natl Acad Sci U S A* 90:5718-5722.
- Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci* 4:324-330.
- Andersen RA (1997) Multimodal integration for the representation of space in the posterior parietal cortex. *Philos Trans R Soc Lond B Biol Sci* 352:1421-1428.
- Andersen RA, Asanuma C, Essick G, Siegel RM (1990) Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *J Comp Neurol* 296:65-113.
- Andersen RA, Snyder LH, Batista AP, Buneo CA, Cohen YE (1998) Posterior parietal areas specialized for eye movements (LIP) and reach (PRR) using a common coordinate frame. *Novartis Found Symp* 218:109-122; discussion 122-108, 171-105.
- Annett J (1966) Training for perceptual skills. *Ergonomics* 9:459-468.
- Apter JT (1946) Eye movements following strychninization of the superior colliculus of cats. In, p 73: *Am Physiological Soc*.
- Avillac M, Ben Hamed S, Duhamel JR (2007) Multisensory integration in the ventral intraparietal area of the macaque monkey. *J Neurosci* 27:1922-1932.
- Bakin JS, Weinberger NM (1996) Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proc Natl Acad Sci U S A* 93:11219-11224.
- Banks MS, Aslin RN, Letson RD (1975) Sensitive period for the development of human binocular vision. *Science* 190:675-677.

- Bao S, Chan VT, Merzenich MM (2001) Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412:79-83.
- Bao S, Chan VT, Zhang LI, Merzenich MM (2003) Suppression of cortical representation through backward conditioning. *Proc Natl Acad Sci U S A* 100:1405-1408.
- Barnes CL, Pandya DN (1992) Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 318:222-244.
- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 17:377-391.
- Bastiaansen M, Hagoort P (2006) Oscillatory neuronal dynamics during language comprehension. *Prog Brain Res* 159:179-196.
- Bavelier D, Neville HJ (2002) Cross-modal plasticity: where and how? *Nat Rev Neurosci* 3:443-452.
- Bavelier D, Brozinsky C, Tomann A, Mitchell T, Neville H, Liu G (2001) Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *J Neurosci* 21:8931-8942.
- Beauchamp MS (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15:145-153.
- Beauchamp MS, Nath AR, Pasalar S (2010) fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci* 30:2414-2417.
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004a) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809-823.
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004b) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190-1192.

- Bende M, Nordin S (1997) Perceptual learning in olfaction: professional wine tasters versus controls. *Physiol Behav* 62:1065-1070.
- Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory--visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp Neurol* 57:849-872.
- Bengoetxea H, Argandona EG, Lafuente JV (2008) Effects of visual experience on vascular endothelial growth factor expression during the postnatal development of the rat visual cortex. *Cereb Cortex* 18:1630-1639.
- Bergan JF, Ro P, Ro D, Knudsen EI (2005) Hunting increases adaptive auditory map plasticity in adult barn owls. *J Neurosci* 25:9816-9820.
- Bertelson P, Radeau M (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys* 29:578-584.
- Bertelson P, Vroomen J, de Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept Psychophys* 62:321-332.
- Bishop CW, Miller LM (2009) A multisensory cortical network for understanding speech in noise. *J Cogn Neurosci* 21:1790-1805.
- Blake R, Sobel KV, James TW (2004) Neural synergy between kinetic vision and touch. *Psychol Sci* 15:397-402.
- Blatt GJ, Andersen RA, Stoner GR (1990) Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *J Comp Neurol* 299:421-445.
- Blau V, van Atteveldt N, Ekkebus M, Goebel R, Blomert L (2009) Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr Biol* 19:503-508.
- Bolognini N, Frassinetti F, Serino A, Ladavas E (2005) "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp Brain Res* 160:273-282.

- Bolognini N, Leo F, Passamonti C, Stein BE, Ladavas E (2007) Multisensory-mediated auditory localization. *Perception* 36:1477-1485.
- Bradley DC, Maxwell M, Andersen RA, Banks MS, Shenoy KV (1996) Mechanisms of heading perception in primate visual cortex. *Science* 273:1544-1547.
- Bremmer F (2005) Navigation in space--the role of the macaque ventral intraparietal area. *J Physiol* 566:29-35.
- Bremmer F, Graf W, Ben Hamed S, Duhamel JR (1999) Eye position encoding in the macaque ventral intraparietal area (VIP). *Neuroreport* 10:873-878.
- Bross M, Borenstein M (1982) Temporal auditory acuity in blind and sighted subjects: a signal detection analysis. *Percept Mot Skills* 55:963-966.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369-384.
- Burnett LR, Stein BE, Chaponis D, Wallace MT (2004) Superior colliculus lesions preferentially disrupt multisensory orientation. *Neuroscience* 124:535-547.
- Burton H, Jones EG (1976) The posterior thalamic region and its cortical projection in New World and Old World monkeys. *J Comp Neurol* 168:249-301.
- Burton H, McLaren DG, Sinclair RJ (2006) Reading embossed capital letters: an fMRI study in blind and sighted individuals. *Hum Brain Mapp* 27:325-339.
- Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300-304.
- Calvert G, Spence C, Stein BE (2004) *The handbook of multisensory processes*. Cambridge, Mass.: MIT Press.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110-1123.

- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649-657.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593-596.
- Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur J Neurosci* 22:2886-2902.
- Cappe C, Rouiller EM, Barone P (2009a) Multisensory anatomical pathways. *Hear Res* 258:28-36.
- Cappe C, Thut G, Romei V, Murray MM (2009b) Selective integration of auditory-visual looming cues by humans. *Neuropsychologia* 47:1045-1052.
- Cardini F, Costantini M, Galati G, Romani GL, Ladavas E, Serino A (2010) Viewing One's Own Face Being Touched Modulates Tactile Perception: An fMRI Study. *J Cogn Neurosci*.
- Carriere BN, Royal DW, Wallace MT (2008) Spatial heterogeneity of cortical receptive fields and its impact on multisensory interactions. *J Neurophysiol* 99:2357-2368.
- Carriere BN, Royal DW, Perrault TJ, Morrison SP, Vaughan JW, Stein BE, Wallace MT (2007) Visual deprivation alters the development of cortical multisensory integration. *J Neurophysiol* 98:2858-2867.
- Casagrande VA, Harting JK, Hall WC, Diamond IT, Martin GF (1972) Superior colliculus of the tree shrew: a structural and functional subdivision into superficial and deep layers. *Science* 177:444-447.
- Chen CM, Lakatos P, Shah AS, Mehta AD, Givre SJ, Javitt DC, Schroeder CE (2007) Functional anatomy and interaction of fast and slow visual pathways in macaque monkeys. *Cereb Cortex* 17:1561-1569.
- Clark B, Graybiel A (1965) Perception of the visual horizontal in normal and labyrinthine defective subjects during prolonged rotation. NSAM-936. *Res Rep U S Nav Sch Aviat Med*:1-7.

- Clark B, Graybiel A (1966) Factors contributing to the delay in the perception of the oculogravic illusion. *Am J Psychol* 79:377-388.
- Colby CL, Duhamel JR, Goldberg ME (1993) Ventral intraparietal area of the macaque: anatomic location and visual response properties. *J Neurophysiol* 69:902-914.
- Collignon O, Voss P, Lassonde M, Lepore F (2009a) Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Exp Brain Res* 192:343-358.
- Collignon O, Davare M, Olivier E, De Volder AG (2009b) Reorganisation of the right occipito-parietal stream for auditory spatial processing in early blind humans. A transcranial magnetic stimulation study. *Brain Topogr* 21:232-240.
- Colonus H, Arndt P (2001) A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys* 63:126-147.
- Colonus H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci* 16:1000-1009.
- Connor S (2000) *Dumbstruck: A cultural history of ventriloquism*: Oxford University Press, USA.
- Conrey B, Pisoni DB (2006) Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J Acoust Soc Am* 119:4065-4073.
- Crist RE, Li W, Gilbert CD (2001) Learning to see: experience and attention in primary visual cortex. *Nat Neurosci* 4:519-525.
- Dahl CD, Logothetis NK, Kayser C (2009) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11924-11932.
- Davis KA, Miller RL, Young ED (1996) Effects of somatosensory and parallel-fiber stimulation on neurons in dorsal cochlear nucleus. *J Neurophysiol* 76:3012-3024.
- de Boer J, Thornton AR (2008) Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J Neurosci* 28:4929-4937.

- de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006a) Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J Comp Neurol* 496:27-71.
- de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006b) Thalamic connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J Comp Neurol* 496:72-96.
- Deco G, Ledberg A, Almeida R, Fuster J (2005) Neural dynamics of cross-modal and cross-temporal associations. *Exp Brain Res* 166:325-336.
- Deshpande G, Hu X, Stilla R, Sathian K (2008) Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. *Neuroimage* 40:1807-1814.
- Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JA (2007) Multisensory integration for timing engages different brain networks. *Neuroimage* 34:764-773.
- Diamond ME, Armstrong-James M, Ebner FF (1993) Experience-dependent plasticity in adult rat barrel cortex. *Proc Natl Acad Sci U S A* 90:2082-2086.
- Diamond ME, Petersen RS, Harris JA (1999) Learning through maps: functional significance of topographic organization in primary sensory cortex. *J Neurobiol* 41:64-68.
- Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9:719-721.
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567-631.
- Drew F (1896) Attention: Experimental and critical. In, pp 533-573: JH Orpha.
- DuBose CN, Cardello AV, Maller O (2006) Effects of colorants and flavorants on identification, perceived flavor intensity, and hedonic quality of fruit-flavored beverages and cake. In, pp 1393-1399: John Wiley & Sons.

- Duhamel JR, Colby CL, Goldberg ME (1998) Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J Neurophysiol* 79:126-136.
- Edwards SB, Ginsburgh CL, Henkel CK, Stein BE (1979) Sources of subcortical projections to the superior colliculus in the cat. *J Comp Neurol* 184:309-329.
- Ehrsson HH, Spence C, Passingham RE (2004) That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305:875-877.
- Engel GR, Dougherty WG (1971) Visual-auditory distance constancy. *Nature* 234:308.
- Fagiolini M, Pizzorusso T, Berardi N, Domenici L, Maffei L (1994) Functional postnatal development of the rat primary visual cortex and the role of visual experience: dark rearing and monocular deprivation. *Vision Res* 34:709-720.
- Fahle M (2004) Perceptual learning: a case for early selection. *J Vis* 4:879-890.
- Fahle M, Edelman S, Poggio T (1995) Fast perceptual learning in hyperacuity. *Vision Res* 35:3003-3013.
- Fajen BR (2007) Rapid recalibration based on optic flow in visually guided action. *Exp Brain Res* 183:61-74.
- Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22:5749-5759.
- Feinsod M, Bach-y-Rita P, Madey JM (1973) Somatosensory evoked responses: latency differences in blind and sighted persons. *Brain Res* 60:219-223.
- Filimon F, Nelson JD, Huang RS, Sereno MI (2009) Multiple parietal reach regions in humans: cortical representations for visual and proprioceptive feedback during on-line reaching. *J Neurosci* 29:2961-2971.
- Fine I, Finney EM, Boynton GM, Dobkins KR (2005) Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex. *J Cogn Neurosci* 17:1621-1637.

- Finney EM, Fine I, Dobkins KR (2001) Visual stimuli activate auditory cortex in the deaf. *Nat Neurosci* 4:1171-1173.
- Fiorentini A, Berardi N (1980) Perceptual learning specific for orientation and spatial frequency. *Nature* 287:43-44.
- Fogassi L, Gallese V, Fadiga L, Luppino G, Matelli M, Rizzolatti G (1996) Coding of peripersonal space in inferior premotor cortex (area F4). *J Neurophysiol* 76:141-157.
- Folta K (2003) Neural fine tuning during Vernier acuity training? *Vision Res* 43:1177-1185.
- Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16:419-423.
- Frens MA, Van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57:802-816.
- Frey RD (1990) Selective attention, event perception and the criterion of acceptability principle: evidence supporting and rejecting the doctrine of prior entry* 1. In, pp 481-530: Elsevier.
- Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9:474-480.
- Friston KJ (1994) Functional and effective connectivity in neuroimaging: a synthesis. In, pp 56-78.
- Friston KJ, Frith CD, Frackowiak RSJ (1993) Time-dependent changes in effective connectivity measured with PET. In, pp 69-79: John Wiley & Sons.
- Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7:773-778.
- Fuster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405:347-351.

- Gabel SF, Misslisch H, Gielen CC, Duysens J (2002) Responses of neurons in area VIP to self-induced and external visual motion. *Exp Brain Res* 147:520-528.
- Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 3:191-197.
- Gauthier I, Curran T, Curby KM, Collins D (2003) Perceptual interference supports a non-modular account of face processing. *Nat Neurosci* 6:428-432.
- Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28:4457-4469.
- Gick B, Derrick D (2009) Aero-tactile integration in speech perception. *Nature* 462:502-504.
- Grant AC, Thiagarajah MC, Sathian K (2000) Tactile perception in blind Braille readers: a psychophysical study of acuity and hyperacuity using gratings and dot patterns. *Percept Psychophys* 62:301-312.
- Grant KW, Greenberg S (2001) Speech intelligibility derived from asynchronous processing of auditory-visual information. In: Citeseer.
- Graziano MS, Yap GS, Gross CG (1994) Coding of visual space by premotor neurons. *Science* 266:1054-1057.
- Graziano MS, Reiss LA, Gross CG (1999) A neuronal representation of the location of nearby sounds. *Nature* 397:428-430.
- Graziano MS, Taylor CS, Moore T (2002) Complex movements evoked by microstimulation of precentral cortex. *Neuron* 34:841-851.
- Grossberg S (1980) How does a brain build a cognitive code? In, pp 1-51.
- Guest S, Catmur C, Lloyd D, Spence C (2002) Audiotactile interactions in roughness perception. *Exp Brain Res* 146:161-171.

- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475-495.
- Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007) Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J Comp Neurol* 502:924-952.
- Hagen MC, Franzen O, McGlone F, Essick G, Dancer C, Pardo JV (2002) Tactile motion activates the human middle temporal/V5 (MT/V5) complex. *Eur J Neurosci* 16:957-964.
- Hagura N, Takei T, Hirose S, Aramaki Y, Matsumura M, Sadato N, Naito E (2007) Activity in the posterior parietal cortex mediates visual dominance over kinesthesia. *J Neurosci* 27:7047-7053.
- Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. *J Cogn Neurosci* 15:20-29.
- Hamlin AJ (1895) On the least observable interval between stimuli addressed to disparate senses and to different organs of the same sense. In, pp 564-575: *JH Orpha*.
- Hanson JV, Heron J, Whitaker D (2008) Recalibration of perceived time across sensory modalities. *Exp Brain Res* 185:347-352.
- Hikosaka K, Iwai E, Saito H, Tanaka K (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *J Neurophysiol* 60:1615-1637.
- Holle H, Obleser J, Rueschemeyer SA, Gunter TC (2010) Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49:875-884.
- Hong B, Acharya S, Thakor N, Gao S (2005) Transient phase synchrony of independent cognitive components underlying scalp EEG. *Conf Proc IEEE Eng Med Biol Soc* 2:2037-2040.
- Horwitz B, Tagamets MA, McIntosh AR (1999) Neural modeling, functional brain imaging, and cognition. *Trends Cogn Sci* 3:91-98.

- Horwitz B, McIntosh AR, Haxby JV, Furey M, Salerno JA, Schapiro MB, Rapoport SI, Grady CL (1995) Network analysis of PET-mapped visual pathways in Alzheimer type dementia. *Neuroreport* 6:2287-2292.
- Hoshino O (2004) Neuronal bases of perceptual learning revealed by a synaptic balance scheme. *Neural Comput* 16:563-594.
- Howard IP, Templeton WB (1966) *Human spatial orientation*. London, New York,: Wiley.
- Hubbard AL, Wilson SM, Callan DE, Dapretto M (2009) Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum Brain Mapp* 30:1028-1037.
- Hubel DH, Wiesel TN (1963) Receptive Fields of Cells in Striate Cortex of Very Young, Visually Inexperienced Kittens. *J Neurophysiol* 26:994-1002.
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215-243.
- Hubel DH, Wiesel TN, LeVay S (1977) Plasticity of ocular dominance columns in monkey striate cortex. *Philos Trans R Soc Lond B Biol Sci* 278:377-409.
- Huerta MF, Harting JK (1984a) The mammalian superior colliculus: studies of its morphology and connections. In, pp 687–773: Plenum Press.
- Huerta MF, Harting JK (1984b) Connectional organization of the superior colliculus. In, pp 286-289: Elsevier.
- Hyvarinen J, Carlson S, Hyvarinen L (1981) Early visual deprivation alters modality of neuronal responses in area 19 of monkey cortex. *Neurosci Lett* 26:239-243.
- Issa NP, Trachtenberg JT, Chapman B, Zahs KR, Stryker MP (1999) The critical period for ocular dominance plasticity in the Ferret's visual cortex. *J Neurosci* 19:6965-6978.

- James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA (2002) Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40:1706-1714.
- Jaskowski P (1993) Selective attention and temporal-order judgment. *Perception* 22:681-689.
- Ji W, Suga N (2003) Development of reorganization of the auditory cortex caused by fear conditioning: effect of atropine. *J Neurophysiol* 90:1904-1909.
- Ji W, Suga N (2008) Tone-specific and nonspecific plasticity of the auditory cortex elicited by pseudoconditioning: role of acetylcholine receptors and the somatosensory cortex. *J Neurophysiol* 100:1384-1396.
- Ji W, Gao E, Suga N (2001) Effects of acetylcholine and atropine on plasticity of central auditory neurons caused by conditioning in bats. *J Neurophysiol* 86:211-225.
- Johnson JS, Newport EL (1989) Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn Psychol* 21:60-99.
- Jones EG, Powell TP (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93:793-820.
- Kaas JH, Merzenich MM, Killackey HP (1983) The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals. *Annu Rev Neurosci* 6:325-356.
- Kadunce DC, Vaughan JW, Wallace MT, Stein BE (2001) The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. *Exp Brain Res* 139:303-310.
- Kang E, Lee DS, Kang H, Hwang CH, Oh SH, Kim CS, Chung JK, Lee MC (2006) The neural correlates of cross-modal interaction in speech perception during a semantic decision task on sentences: a PET study. *Neuroimage* 32:423-431.
- Karni A, Sagi D (1993) The time course of learning a visual skill. *Nature* 365:250-252.

- Kavounoudias A, Roll JP, Anton JL, Nazarian B, Roth M, Roll R (2008) Proprio-tactile integration for kinesthetic perception: an fMRI study. *Neuropsychologia* 46:567-575.
- Kayser C, Logothetis NK (2009) Directed Interactions Between Auditory and Superior Temporal Cortices and their Role in Sensory Integration. *Front Integr Neurosci* 3:7.
- Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. *Neuron* 48:373-384.
- Keetels M, Vroomen J (2005) The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Exp Brain Res* 167:635-640.
- Kellogg WN (1962) Sonar system of the blind. *Science* 137:399-404.
- Kilgard M (2003) Cholinergic modulation of skill learning and plasticity. In, pp 678-680: Elsevier.
- Kilgard MP, Merzenich MM (1998a) Cortical map reorganization enabled by nucleus basalis activity. *Science* 279:1714-1718.
- Kilgard MP, Merzenich MM (1998b) Plasticity of temporal information processing in the primary auditory cortex. *Nat Neurosci* 1:727-731.
- Kilgard MP, Merzenich MM (2002) Order-sensitive plasticity in adult primary auditory cortex. *Proc Natl Acad Sci U S A* 99:3205-3209.
- Kim RS, Seitz AR, Shams L (2008) Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One* 3:e1532.
- Koenig T, Studer D, Hubl D, Melie L, Strik WK (2005) Brain connectivity at different time-scales measured with EEG. *Philos Trans R Soc Lond B Biol Sci* 360:1015-1023.
- Kopinska A, Harris LR (2004) Simultaneity constancy. *Perception* 33:1049-1060.

- Korte M, Rauschecker JP (1993) Auditory spatial tuning of cortical neurons is sharpened in cats with early blindness. *J Neurophysiol* 70:1717-1721.
- Krueger J, Royal DW, Fister MC, Wallace MT (2009) Spatial receptive field organization of multisensory neurons and its impact on multisensory interactions. *Hear Res* 258:47-54.
- Kujala T, Alho K, Huotilainen M, Ilmoniemi RJ, Lehtokoski A, Leinonen A, Rinne T, Salonen O, Sinkkonen J, Standertskjold-Nordenstam CG, Naatanen R (1997) Electrophysiological evidence for cross-modal plasticity in humans with early- and late-onset blindness. *Psychophysiology* 34:213-216.
- Lacey S, Tal N, Amedi A, Sathian K (2009) A putative model of multisensory object representation. *Brain Topogr* 21:269-274.
- Ladavas E, Farne A (2004) Visuo-tactile representation of near-the-body space. *J Physiol Paris* 98:161-170.
- Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron* 53:279-292.
- Lakatos P, Pincze Z, Fu KM, Javitt DC, Karmos G, Schroeder CE (2005) Timing of pure tone and noise-evoked responses in macaque auditory cortex. *Neuroreport* 16:933-937.
- Law CT, Gold JI (2008) Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat Neurosci* 11:505-513.
- Lee L, Friston K, Horwitz B (2006) Large-scale neural models and dynamic causal modelling. *Neuroimage* 30:1243-1254.
- Leo F, Bertini C, di Pellegrino G, Ladavas E (2008) Multisensory integration for orienting responses in humans requires the activation of the superior colliculus. *Exp Brain Res* 186:67-77.
- Leon MI, Shadlen MN (2003) Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* 38:317-327.

- Lewald J, Ehrenstein WH, Guski R (2001) Spatio-temporal constraints for auditory--visual integration. *Behav Brain Res* 121:69-79.
- Lewald J, Foltys H, Topper R (2002) Role of the posterior parietal cortex in spatial hearing. *J Neurosci* 22:RC207.
- Lewis JW, Van Essen DC (2000) Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol* 428:112-137.
- Linkenhoker BA, Knudsen EI (2002) Incremental training increases the plasticity of the auditory space map in adult barn owls. *Nature* 419:293-296.
- Liu Z (1999) Perceptual learning in motion discrimination that generalizes across motion directions. *Proc Natl Acad Sci U S A* 96:14085-14087.
- Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC (2009) Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One* 4:e4638.
- Macaluso E, George N, Dolan R, Spence C, Driver J (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage* 21:725-732.
- MacCallum RC, Austin JT (2000) Applications of structural equation modeling in psychological research. *Annu Rev Psychol* 51:201-226.
- MacLeod A, Summerfield Q (1987) Quantifying the contribution of vision to speech perception in noise. *Br J Audiol* 21:131-141.
- Massaro DW, Cohen MM (1993) Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables* 1. In, pp 127-134: Elsevier.
- Massaro DW, Cohen MM, Smeele PM (1996) Perception of asynchronous and conflicting visual and auditory speech. *J Acoust Soc Am* 100:1777-1786.
- Masuda N (2009) Selective population rate coding: a possible computational role of gamma oscillations in selective attention. *Neural Comput* 21:3335-3362.

- Maunsell JH, van Essen DC (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurosci* 3:2563-2586.
- Mazzoni P, Bracewell RM, Barash S, Andersen RA (1996) Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *J Neurophysiol* 75:1233-1241.
- McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. *Nat Neurosci* 8:1197-1202.
- McGrath M, Summerfield Q (1985) Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J Acoust Soc Am* 77:678-685.
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746-748.
- McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1994) Network analysis of cortical visual pathways mapped with PET. *J Neurosci* 14:655-666.
- Meienbrock A, Naumer MJ, Doehrmann O, Singer W, Muckli L (2007) Retinotopic effects during spatial audio-visual integration. *Neuropsychologia* 45:531-539.
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389-391.
- Meredith MA, Stein BE (1985) Descending efferents from the superior colliculus relay integrated multisensory information. *Science* 227:657-659.
- Meredith MA, Stein BE (1986a) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol* 56:640-662.
- Meredith MA, Stein BE (1986b) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365:350-354.

- Meredith MA, Clemo HR (1989) Auditory cortical projection from the anterior ectosylvian sulcus (Field AES) to the superior colliculus in the cat: an anatomical and electrophysiological study. *J Comp Neurol* 289:687-707.
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215-3229.
- Merzenich MM, Kaas JH, Wall J, Nelson RJ, Sur M, Felleman D (1983) Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience* 8:33-55.
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cognit Psychol* 14:247-279.
- Miller J (1986) Timecourse of coactivation in bimodal divided attention. *Percept Psychophys* 40:331-343.
- Miquee A, Xerri C, Rainville C, Anton JL, Nazarian B, Roth M, Zennou-Azogui Y (2008) Neuronal substrates of haptic shape encoding and matching: a functional magnetic resonance imaging study. *Neuroscience* 152:29-39.
- Miyazaki M, Nozaki D, Nakajima Y (2005) Testing Bayesian models of human coincidence timing. *J Neurophysiol* 94:395-399.
- Miyazaki M, Yamamoto S, Uchida S, Kitazawa S (2006) Bayesian calibration of simultaneity in tactile temporal order judgment. *Nat Neurosci* 9:875-877.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115-128.
- Morrell F (1972) Visual system's view of acoustic space. *Nature* 238:44-46.
- Mossbridge JA, Scissors BN, Wright BA (2008) Learning and generalization on asynchrony and order tasks at sound offset: implications for underlying neural circuitry. *Learn Mem* 15:13-20.

- Mountcastle VB, Lynch JC, Georgopoulos A, Sakata H, Acuna C (1975) Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *J Neurophysiol* 38:871-908.
- Muchnik C, Efrati M, Nemeth E, Malin M, Hildesheimer M (1991) Central auditory skills in blind and sighted subjects. *Scand Audiol* 20:19-23.
- Mullette-Gillman OA, Cohen YE, Groh JM (2005) Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *J Neurophysiol* 94:2331-2352.
- Munhall KG, Gribble P, Sacco L, Ward M (1996) Temporal constraints on the McGurk effect. *Percept Psychophys* 58:351-362.
- Musacchia G, Schroeder CE (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear Res* 258:72-79.
- Nagarajan SS, Blake DT, Wright BA, Byl N, Merzenich MM (1998) Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J Neurosci* 18:1559-1570.
- Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Res Cogn Brain Res* 25:499-507.
- Neville HJ, Schmidt A, Kutas M (1983) Altered visual-evoked potentials in congenitally deaf adults. *Brain Res* 266:127-132.
- Niemeyer W, Starlinger I (1981) Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. II. Central functions. *Audiology* 20:510-515.
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431-11441.

- Ogasawara H, Doi T, Kawato M (2008) Systems biology perspectives on cerebellar long-term depression. *Neurosignals* 16:300-317.
- Pandey PC, Kunov H, Abel SM (1986) Disruptive effects of auditory signal delay on speech perception with lipreading. *J Aud Res* 26:27-41.
- Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M (1998) Increased auditory cortical representation in musicians. *Nature* 392:811-814.
- Pascual-Leone A, Torres F (1993) Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. *Brain* 116 (Pt 1):39-52.
- Pleger B, Dinse HR, Ragert P, Schwenkreis P, Malin JP, Tegenthoff M (2001) Shifts in cortical representations predict human discrimination improvement. *Proc Natl Acad Sci U S A* 98:12255-12260.
- Poggio T, Fahle M, Edelman S (1992) Fast perceptual learning in visual hyperacuity. *Science* 256:1018-1021.
- Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26:4970-4982.
- Polley DB, Heiser MA, Blake DT, Schreiner CE, Merzenich MM (2004) Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. *Proc Natl Acad Sci U S A* 101:16351-16356.
- Polley DB, Hillock AR, Spankovich C, Popescu MV, Royal DW, Wallace MT (2008) Development and plasticity of intra- and intersensory information processing. *J Am Acad Audiol* 19:780-798.
- Popescu MV, Polley DB (2010) Monaural deprivation disrupts development of binaural selectivity in auditory midbrain and cortex. *Neuron* 65:718-731.
- Ptito M, Matteau I, Gjedde A, Kupers R (2009) Recruitment of the middle temporal area by tactile motion in congenital blindness. *Neuroreport* 20:543-547.

- Rammsayer TH, Vogel WH (1992) Pharmacologic properties of the internal clock underlying time perception in humans. *Neuropsychobiology* 26:71-80.
- Rauschecker JP, Harris LR (1983) Auditory compensation of the effects of visual deprivation in the cat's superior colliculus. *Exp Brain Res* 50:69-83.
- Rauschecker JP, Korte M (1993) Auditory compensation for early blindness in cat cerebral cortex. *J Neurosci* 13:4538-4548.
- Rice CE, Feinstein SH, Schusterman RJ (1965) Echo-Detection Ability of the Blind: Size and Distance Factors. *J Exp Psychol* 70:246-255.
- Rockland KS, Pandya DN (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res* 179:3-20.
- Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *Int J Psychophysiol* 50:19-26.
- Roder B, Wallace M (2010) Development and plasticity of multisensory functions. *Restor Neurol Neurosci* 28:141-142.
- Rolls BJ, Rowe EA, Rolls ET (1982) How sensory properties of foods affect human feeding behavior. In, pp 409-417: Elsevier.
- Romanski LM (2007) Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb Cortex* 17 Suppl 1:i61-69.
- Royal DW, Carriere BN, Wallace MT (2009) Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. *Exp Brain Res* 198:127-136.
- Royal DW, Krueger J, Fister MC, Wallace MT (2010) Adult plasticity of spatiotemporal receptive fields of multisensory superior colliculus neurons following early visual deprivation. *Restor Neurol Neurosci* 28:259-270.
- Sadato N, Pascual-Leone A, Grafman J, Ibanez V, Deiber MP, Dold G, Hallett M (1996) Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* 380:526-528.

- Samson FK, Clarey JC, Barone P, Imig TJ (1993) Effects of ear plugging on single-unit azimuth sensitivity in cat primary auditory cortex. I. Evidence for monaural directional cues. *J Neurophysiol* 70:492-511.
- Schaefer M, Heinze HJ, Rotte M (2009) My third arm: shifts in topography of the somatosensory homunculus predict feeling of an artificial supernumerary arm. *Hum Brain Mapp* 30:1413-1420.
- Schaefer M, Flor H, Heinze HJ, Rotte M (2006) Dynamic modulation of the primary somatosensory cortex during seeing and feeling a touched hand. *Neuroimage* 29:587-592.
- Schlack A, Hoffmann KP, Bremmer F (2002) Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16:1877-1886.
- Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann KP, Bremmer F (2005) Multisensory space representations in the macaque ventral intraparietal area. *J Neurosci* 25:4616-4625.
- Schneider KA, Bavelier D (2003) Components of visual prior entry. *Cogn Psychol* 47:333-366.
- Schneider TR, Engel AK, Debener S (2008) Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp Psychol* 55:121-132.
- Schoups A, Vogels R, Qian N, Orban G (2001) Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412:549-553.
- Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res* 14:187-198.
- Schwartz S, Maquet P, Frith C (2002) Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proc Natl Acad Sci U S A* 99:17137-17142.
- Seitz AR, Kim R, Shams L (2006) Sound facilitates visual learning. *Curr Biol* 16:1422-1427.

- Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. *Nature* 385:308.
- Seltzer B, Pandya DN (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol* 343:445-463.
- Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45:561-571.
- Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. *Nature* 408:788.
- Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14:147-152.
- Sherman SM, Guillery RW (2002) The role of the thalamus in the flow of information to the cortex. *Philos Trans R Soc Lond B Biol Sci* 357:1695-1708.
- Shiu LP, Pashler H (1992) Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept Psychophys* 52:582-588.
- Shore DI, Spence C, Klein RM (2001) Visual prior entry. *Psychol Sci* 12:205-212.
- Shore SE (2005) Multisensory integration in the dorsal cochlear nucleus: unit responses to acoustic and trigeminal ganglion stimulation. *Eur J Neurosci* 21:3334-3348.
- Simons DJ, Land PW (1987) Early experience of tactile stimulation influences organization of somatic sensory cortex. *Nature* 326:694-697.
- Skrandies W, Jedynek A, Fahle M (2001) Perceptual learning: psychophysical thresholds and electrical brain topography. *Int J Psychophysiol* 41:119-129.
- Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12:7-10.

- Snyder LH, Batista AP, Andersen RA (2000) Saccade-related activity in the parietal reach region. *J Neurophysiol* 83:1099-1102.
- Soto-Faraco S, Spence C, Kingstone A (2004) Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *J Exp Psychol Hum Percept Perform* 30:330-345.
- Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. *J Exp Psychol Gen* 130:799-832.
- Spengler F, Roberts TP, Poeppel D, Byl N, Wang X, Rowley HA, Merzenich MM (1997) Learning transfer and neuronal plasticity in humans trained in tactile discrimination. *Neurosci Lett* 232:151-154.
- Starlinger I, Niemeier W (1981) Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. I. Peripheral functions. *Audiology* 20:503-509.
- Stein BE, Arigbede MO (1972) Unimodal and multimodal response properties of neurons in the cat's superior colliculus. *Exp Neurol* 36:179-196.
- Stein BE, Meredith MA (1993) *The merging of the senses*. Cambridge, Mass.: MIT Press.
- Stein BE, Wallace MT (1996) Comparisons of cross-modality integration in midbrain and cortex. *Prog Brain Res* 112:289-299.
- Stein BE, Spencer RF, Edwards SB (1983) Corticotectal and corticothalamic efferent projections of SIV somatosensory cortex in cat. *J Neurophysiol* 50:896-909.
- Stein BE, Huneycutt WS, Meredith MA (1988) Neurons and behavior: the same rules of multisensory integration apply. *Brain Res* 448:355-358.
- Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. In, pp 12-24: MIT Press.

- Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38:387-401.
- Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ (2010) Ten simple rules for dynamic causal modeling. *Neuroimage* 49:3099-3109.
- Stetson C, Cui X, Montague PR, Eagleman DM (2006) Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron* 51:651-659.
- Stevens HC (1904) A Simple Complication Pendulum for Qualitative Work. In, pp 581-581: Louis N. Wilson.
- Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44:1210-1223.
- Stevenson RA, Kim S, James TW (2009) An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Exp Brain Res* 198:183-194.
- Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of asynchronous audiovisual speech perception. *Neuroimage* 49:3308-3318.
- Stevenson RJ (2001) Perceptual learning with odors: implications for psychological accounts of odor quality perception. *Psychon Bull Rev* 8:708-712.
- Stone JV, Hunkin NM, Porrill J, Wood R, Keeler V, Beanland M, Port M, Porter NR (2001) When is now? Perception of simultaneity. *Proc R Soc Lond B Biol Sci* 268:31-38.
- Stricanne B, Andersen RA, Mazzoni P (1996) Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J Neurophysiol* 76:2071-2076.
- Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26:11138-11147.

- Sumby WH, Pollack I (1954) Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America* 26:212.
- Tanaka E, Kida T, Inui K, Kakigi R (2009) Change-driven cortical activation in multisensory environments: an MEG study. *Neuroimage* 48:464-474.
- Taylor JL, McCloskey DI (1991) Illusions of head and visual target displacement induced by vibration of neck muscles. *Brain* 114 (Pt 2):755-759.
- Tortelly A, Reinoso-Suarez F, Llamas A (1980) Projections from non-visual cortical areas to the superior colliculus demonstrated by retrograde transport of HRP in the cat. *Brain Res* 188:543-549.
- Vallar G, Ronchi R (2009) Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Exp Brain Res* 192:533-551.
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271-282.
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962-974.
- Van Boven RW, Hamilton RH, Kauffman T, Keenan JP, Pascual-Leone A (2000) Tactile spatial resolution in blind braille readers. *Neurology* 54:2230-2236.
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102:1181-1186.
- van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598-607.
- Vibell J, Klinge C, Zampini M, Spence C, Nobre AC (2007) Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *J Cogn Neurosci* 19:109-120.

- von Kriegstein K, Giraud AL (2006) Implicit multisensory associations influence voice recognition. *PLoS Biol* 4:e326.
- Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res* 22:32-35.
- Wallace MT, Stein BE (1994) Cross-modal synthesis in the midbrain depends on input from cortex. *J Neurophysiol* 71:429-432.
- Wallace MT, Stein BE (1996) Sensory organization of the superior colliculus in cat and monkey. *Prog Brain Res* 112:301-311.
- Wallace MT, Stein BE (2007) Early experience determines how the senses will interact. *J Neurophysiol* 97:921-926.
- Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91:484-488.
- Wallace MT, Meredith MA, Stein BE (1993) Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol* 69:1797-1809.
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246-1266.
- Wallace MT, Perrault TJ, Jr., Hairston WD, Stein BE (2004) Visual experience is necessary for the development of multisensory integration. *J Neurosci* 24:9580-9584.
- Wallace W, Bear MF (2004) A morphological correlate of synaptic scaling in visual cortex. *J Neurosci* 24:6928-6938.
- Watanabe T, Nanez JE, Sr., Koyama S, Mukai I, Liederman J, Sasaki Y (2002) Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nat Neurosci* 5:1003-1009.

- Watkins S, Shams L, Josephs O, Rees G (2007) Activity in human V1 follows multisensory perception. *Neuroimage* 37:572-578.
- Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31:1247-1256.
- Weinberger NM (2004) Specific long-term memory traces in primary auditory cortex. *Nat Rev Neurosci* 5:279-290.
- Werner S, Noppeney U (2009) Superadditive Responses in Superior Temporal Sulcus Predict Audiovisual Benefits in Object Categorization. *Cereb Cortex*.
- Wiesel TN, Hubel DH (1965) Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J Neurophysiol* 28:1029-1040.
- Wilson DA, Stevenson RJ (2003) Olfactory perceptual learning: the critical role of memory in odor discrimination. *Neurosci Biobehav Rev* 27:307-328.
- Wright BA, Buonomano DV, Mahncke HW, Merzenich MM (1997) Learning and generalization of auditory temporal-interval discrimination in humans. *J Neurosci* 17:3956-3963.
- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034-1043.
- Xiao LQ, Zhang JY, Wang R, Klein SA, Levi DM, Yu C (2008) Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr Biol* 18:1922-1926.
- Young ED, Nelken I, Conley RA (1995) Somatosensory effects on neurons in dorsal cochlear nucleus. *J Neurophysiol* 73:743-765.
- Zampini M, Spence C (2005) The role of auditory cues in modulating the perceived crispness and staleness of potato chips. In, pp 347-363: John Wiley & Sons.

- Zampini M, Shore DI, Spence C (2003) Audiovisual temporal order judgments. *Exp Brain Res* 152:198-210.
- Zampini M, Shore DI, Spence C (2005a) Audiovisual prior entry. *Neurosci Lett* 381:217-222.
- Zampini M, Guest S, Shore DI, Spence C (2005b) Audio-visual simultaneity judgments. *Percept Psychophys* 67:531-544.
- Zhang LI, Bao S, Merzenich MM (2001) Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nat Neurosci* 4:1123-1130.
- Zhang T, Xiao LQ, Klein SA, Levi DM, Yu C (2010) Decoupling location specificity from perceptual learning of orientation discrimination. *Vision Res* 50:368-374.
- Zhou YD, Fuster JM (2000) Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proc Natl Acad Sci U S A* 97:9777-9782.
- Zhou YD, Fuster JM (2004) Somatosensory cell response to an auditory cue in a haptic memory task. *Behav Brain Res* 153:573-578.

CHAPTER II

PERCEPTUAL TRAINING NARROWS

THE TEMPORAL WINDOW OF MULTISENSORY BINDING[†]

Abstract

The brain's ability to bind incoming auditory and visual stimuli depends critically on the temporal structure of this information. Specifically, there exists a temporal window of audiovisual integration within which stimuli are highly likely to be bound together and perceived as part of the same environmental event. Several studies have described the temporal bounds of this window, but few have investigated its malleability. Here, the plasticity in the size of this temporal window was investigated using a perceptual learning paradigm in which participants were given feedback during a two-alternative forced-choice (2-AFC) audiovisual simultaneity judgment task. Training resulted in a marked (i.e., approximately 40%) narrowing in the size of the window. To rule out the possibility that this narrowing was the result of changes in cognitive biases, a second experiment employing a two-interval forced choice (2-IFC) paradigm was undertaken during which participants were instructed to identify a simultaneously-presented audiovisual pair presented within one of two intervals. The 2-IFC paradigm resulted in a narrowing that was similar in both degree and dynamics to that using the 2-AFC approach. Together, these results illustrate that different methods of multisensory

[†] The contents of this chapter were first published as:
Albert R. Powers, III, Andrea R. Hillock, and Mark T. Wallace **Perceptual Training Narrows the Temporal Window of Multisensory Binding** *J. Neurosci.* 29: 12265-12274;
doi:10.1523/JNEUROSCI.3501-09.2009

perceptual training can result in substantial alterations in the circuits underlying the perception of audiovisual simultaneity. These findings suggest a high degree of flexibility in multisensory temporal processing and have important implications for interventional strategies that may be used to ameliorate clinical conditions (e.g., autism, dyslexia) in which multisensory temporal function may be impaired.

Introduction

The proper integration of information from the different sensory modalities is central to our ability to perceive the world in an accurate and meaningful way. One of the most formidable tasks the brain faces in this process comes in determining whether stimuli from different modalities were generated by a single external source or come from different sources. It is not surprising, then, that one of the key cues in this likelihood determination is spatial location (Meredith and Stein, 1986; Wallace et al., 1992; Teder-Salejarvi et al., 2005), since stimuli that are spatially proximate are likely to be associated with a common event, and stimuli that are spatially disparate are unlikely to be of common origin. Similarly, the temporal structure of a multisensory stimulus pair provides important probabilistic information as to the sources of sensory information. However, given the differing propagation times for environmental energies in each of the sensory systems, the temporal relationship of a stimulus pair derived from the same event must be flexibly specified. Consequently, the concept of a multisensory temporal binding window emerges as a useful construct. Within this window, the combination of information from two modalities results in significant changes in neural, behavioral and perceptual responses (Dixon and Spitz, 1980; McGrath and Summerfield, 1985; Meredith

et al., 1987a; Lewkowicz, 1996; Colonius and Diederich, 2004). Multisensory temporal processes have been best examined in the audiovisual domain and have capitalized on tools such as simultaneity judgment tasks to define the important time scales for audiovisual binding.

Although developmental studies have highlighted that significant changes take place in multisensory temporal processing as maturation progresses (Lewkowicz, 1996; Wallace et al., 1997; Wallace and Stein, 1997; Lewkowicz and Ghazanfar, 2006; Lewkowicz et al., 2008), few have looked at the possible malleability of these processes in the adult. Those that have examined the window's flexibility focused on changes in point measures such as the PSS and have shown that repeated exposure to asynchronous multisensory combinations biases judgments in the direction of the repeated exposure (Fujisaki et al., 2004; Vroomen et al., 2004; Navarra et al., 2005; Vatakis et al., 2007; Hanson et al., 2008; Keetels and Vroomen, 2008). In contrast, no work has examined whether the size of the multisensory temporal window can be enlarged or contracted, a change that would be of strong ethological and perceptual relevance because of the importance of this window in the binding of cross-modal cues and because there is increasing evidence that this window may be enlarged in several prominent neurobiological disorders (de Gelder et al., 2003; Virsu et al., 2003; de Gelder et al., 2005; Hairston et al., 2005). In the current study, we set out to examine whether we could alter the temporal characteristics of multisensory processing in adults by engaging participants in two perceptual training paradigms in which they were given feedback as to the correctness of their simultaneity judgments.

Materials and Methods

2-AFC Training

Subjects

Twenty-two (22) Vanderbilt undergraduate and graduate students (mean age 20.73; 11 female) underwent the 2-alternative forced-choice (2-AFC) training portion of the study. All participants had self-reported normal hearing and vision, and none had any personal or close family history of neurological or psychiatric disorders. All recruitment and experimental procedures were approved by the Vanderbilt University Institutional Review Board (IRB).

2-AFC Simultaneity Judgment Assessment

In this task (**Fig 2.1**), participants judged whether the occurrence of a visual stimulus and an auditory stimulus were 'simultaneous' or 'non-simultaneous' by pressing 1 or 2, respectively, on a response box (Psychology Software Tools Response Box Model 200A). Participants were seated in a dark and quiet room 48 cm from a computer monitor. E-Prime 2.0 (2.0.1.109) was used to control all experiments.

A white crosshair fixation marker (1 cm x 1 cm) on a black background appeared 1 second before the stimuli were presented and persisted throughout the duration of each trial. The visual stimulus consisted of a white ring on a black background that subtended

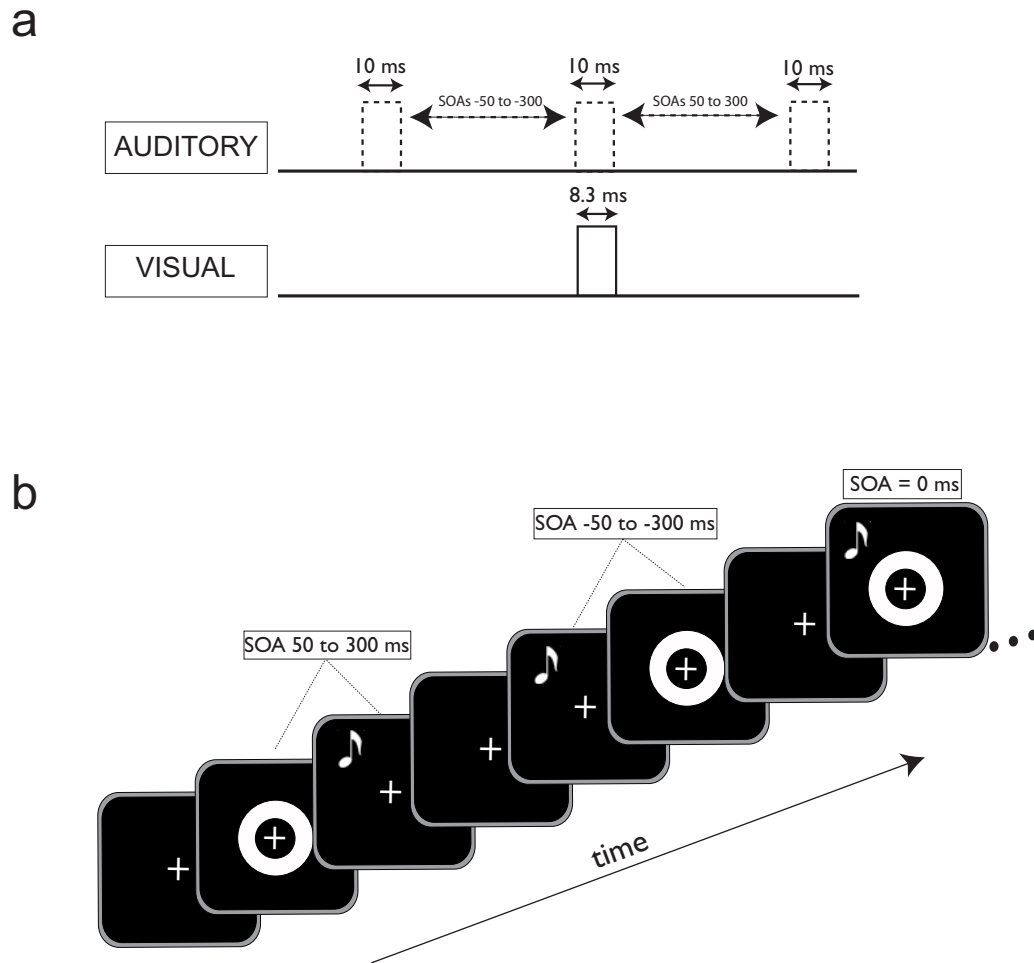


Figure 2.1. Simultaneity judgment protocol **a.** Temporal relationship between visual (ring flash) and auditory (tone pip) stimuli used in both the 2-AFC and 2-IFC experiments. A tone pip was paired with a single ring flash presented at a stimulus-onset asynchrony ranging from -300 (auditory preceding visual) to 300 ms (visual preceding auditory, in 50 ms steps) for each trial. The duration of the visual and auditory stimuli was 11 ms. **b.** Representation of a portion of a trial sequence for the 2-AFC task. Participants were instructed to maintain fixation on the central cross for the duration of the trial and respond after presentation of each pair. In the 2-IFC version of the task, participants were instructed to respond after every second presentation to determine which of the preceding two contained the simultaneous audiovisual pair.

15° of visual space with an outer diameter of 12.4 cm and an inner diameter of 6.0 cm (area = 369.8 cm²). This stimulus was presented for one refresh cycle on a high refresh-rate monitor (NEC MultiSync FE992, 120 Hz) and hence were 8.3 ms in duration.

The auditory stimulus was a 10-ms, 1800-Hz tone burst presented to both ears via headphones (Phillips SBC HN110) with no interaural time or level differences. The acoustic stimulus was calibrated with a Larson-Davis sound level meter (Model # 814). Auditory stimuli were presented at 110.4 dB SPL unweighted using impulse detection and flat weighting settings.

The stimuli had stimulus onset asynchronies (SOAs) ranging from -300 ms (auditory stimulus leading) to 300 ms (visual stimulus leading) at 50 ms intervals. The timing of all stimuli was verified externally with an oscilloscope within an error tolerance of 10 ms arising from the inherent timing error of the auditory presentation hardware and drivers. In the Simultaneity Judgment Assessment task, the lags were equally distributed. A total of 325 trials made up the task (25 cycles x 13 trials/cycle).

2-AFC Simultaneity Judgment Training

The training tasks differed from assessments in that after making a response, the subject was presented with either the phrase “Correct!” paired with a happy face, or “Incorrect” paired with a sad face corresponding to the correctness of their choice. These faces (area = 37.4 cm², happy = yellow, sad = blue) were presented in the center of the screen for 500 ms. The white ring and fixation were the same size as in assessment trials and were presented for the same amount of time. Only SOAs between -150 and 150 ms, broken into 50 ms intervals, were used for the training phase. In addition, in the training

phase the SOAs were not equally distributed: the veridical simultaneous condition had a 6:1 ratio to any of the other 6 non-simultaneous conditions. In this way there was an equal likelihood of simultaneous/non-simultaneous conditions, minimizing concerns about response bias. There were 120 trials in the training phase (20 cycles x 6 trials/cycle). See **Figs. 2.1a** and **2.1b** for illustrations of the temporal structure of stimulus presentation.

2-AFC Training Protocol

Training consisted of 5 hours (1 hour per day) during which participants took part first in a pre-training simultaneity judgment assessment, then in 3 shorter simultaneity judgment training blocks, followed by a post-training simultaneity judgment assessment. An additional baseline assessment was performed at the outset of the study for each subject, followed by the typical training day; this served to detect any practice effects that may have resulted from completion of the assessment itself.

Follow-Up Assessment

After one week without training, a subset of training subjects (n=16) returned to the lab and underwent one simultaneity judgment assessment without any training.

2-AFC Exposure

Subjects

Fourteen (14) Vanderbilt undergraduate and graduate students (mean age 19.50; 4 female) underwent the 2-alternative forced-choice (2-AFC) exposure portion of the study.

All participants had self-reported normal hearing and vision, and none had any personal or close family history of neurological or psychiatric disorders.

Exposure Protocol

The exposure portion of the study differed from the 2-AFC training protocol only in that in lieu of the training blocks, participants underwent 2-AFC exposure blocks of the same length. Thus, all participants in both Experiment 1 and Experiment 2 took part in the same number of 2-AFC simultaneity judgment assessments. The details of the exposure blocks are outlined below.

2-AFC Exposure

To maintain attention, the 2-AFC exposure blocks consisted of an oddball task wherein participants were exposed to the same ring-tone pairs present in the simultaneity judgment training sessions but were instructed to press a button when they saw a red ring. As in the simultaneity judgment training sessions, SOAs were not equally distributed: the veridical simultaneous condition had a 6:1 ratio to any of the other 6 non-simultaneous conditions. Red rings occurred with the same probability across all conditions, and were 1/10 as likely to appear as white rings. The rings and fixation were the same size as in the assessment trial and were presented for the same amount of time; the tone was identical to that presented during assessment and training sessions. Only SOAs between -150 and 150 ms, in steps of 50-ms intervals, were used for this task.

2-IFC Training

Subjects

Twenty (20) Vanderbilt undergraduate and graduate students (mean age 20.20; 13 female) underwent the 2-interval forced-choice (2-IFC) training portion of the study. All participants had self-reported normal hearing and vision, and none had any personal or close family history of neurological or psychiatric disorders.

2-IFC Simultaneity Judgment Assessment

The 2-IFC simultaneity judgment assessment employed exactly the same stimuli as those used in the 2-AFC task. In this task, however, participants were presented with two visual-auditory pairs, one with an SOA of zero (simultaneously-presented) and one with a non-zero SOA (non-simultaneously presented). Presentations were separated by 1 second, during which a fixation cross alone was presented. Instructions asked participants to indicate by button-press which interval (first or second presentation) contained the flash and beep that happened at the same time. Participants were instructed to respond as quickly as possible. Simultaneous pairings were as likely to be presented in the first interval as in the second. A simultaneous-simultaneous condition was present in equal representation to other SOAs as a catch trial.

2-IFC Simultaneity Judgment Training

The training phase of the 2-IFC portion of the study was identical to that of the assessment phase with two exceptions: 1) participants were given feedback as to the accuracy of their responses after each trial, in the same manner described in the 2-AFC

training; 2) in a manner similar to the 2-AFC simultaneity judgment training protocol, the range of SOAs presented during training (-150 ms to 150 ms by 50-ms increments) was restricted in training as compared to assessment (-300 ms to 300 ms). However, unlike the 2-AFC version of this training, and by virtue of the 2-IFC structure, the ratio of simultaneous to non-simultaneous presentation was always 1:1.

2-IFC Training Protocol

Participants underwent training in five 1-hour blocks (one hour per day) on the two-interval forced-choice version of the simultaneity judgment task. Similar to the 2-AFC training protocol, each day's 2-IFC training began with a simultaneity judgment assessment followed by three shorter blocks of training, and ended with a post-training simultaneity judgment assessment.

Follow-Up Assessment

A subset of training subjects (n=9) returned to the lab one week after cessation of training and underwent one simultaneity judgment assessment without any training.

Data Analysis

All data were imported from E-Prime 2.0 .txt files into MatLab 7.7.0.471 R2008b (The Mathworks, Inc., Cambridge, MA) via a custom-made script for this purpose. Individual subject raw data were used to calculate the mean probability of simultaneity judgment (2-AFC) and accuracy (2-IFC) at each SOA for all assessments. These means were then analyzed in multiple ways as summarized in the following sections.

Grand Mean SOA Analysis

To determine how overall group probability of simultaneity judgment (2-AFC) or accuracy (2-IFC) changed after training or exposure, individual means at each SOA were averaged to produce the grand average plots shown in Figures 2b, 3b, 4a, 5b, and 6a. Statistical analysis included performance of a two-factor (group, SOA) repeated-measures analysis of variance (ANOVA), followed by post-hoc t-tests (with Holm correction for multiple comparisons) if significant to determine which SOAs showed statistically significant variation from pre-training to post-training assessment.

Window Size Estimation

Individual mean data were fit with two sigmoid curves generated using the MatLab *glmfit* function, splitting the data into left (auditory presented first) and right (visual presented first) sides and fitting them separately. A criterion at which to measure each individual temporal window size was then established. For the 2-AFC tasks, this criterion was equal to 75% of the maximum data point at baseline assessment. For the 2-IFC task, this criterion was set at half the distance between individuals' lowest accuracy point at baseline assessment and 1 (also about 75% accuracy). These criterion lines were then used to assess the width of the distributions produced by each individual's assessment data throughout the duration of the training period. Distribution width was then assessed for both the left side (from zero to the left-most point at which the sigmoid curve crossed the criterion line) and the right side (from zero to right intersection point) and then combined to get an estimation of total distribution width. This was then used as a proxy for the size of each individual's window at each assessment. An example of the

result of this process may be seen in **Figure 2.2b**. It should be noted that, when mean data from any individual assessment were unable to be fit with a sigmoid curve, all data from this individual were discarded for analysis of window size progression. Group-level analysis of differences in window size across time was conducted by performing a repeated-measures ANOVA (within-subject factor, assessment number) followed by post-hoc t-tests (corrected via the Holm method for multiple comparisons) to determine which differences between assessment measures were responsible for the variance observed.

Results

2-AFC

Judgments of audiovisual simultaneity can be used to define a multisensory temporal binding window

The data produced by the 2-AFC training protocol from one participant are shown in **Figure 2.2a**. Here, the mean probability of simultaneity judgment is plotted as a function of SOA and then fitted with two sigmoid curves to model the left and right sides of the plot. The resulting distribution was used to create a singular metric to serve as an index of the multisensory temporal binding window. The value of this window was set as the breadth of the distribution (in ms) at which individual participants reported simultaneity at or over 75% of their maximum at baseline (full width at 75% height). This level was chosen because it represents half the distance between the 50% level and each individual's highest likelihood of reporting simultaneity. Note that for this

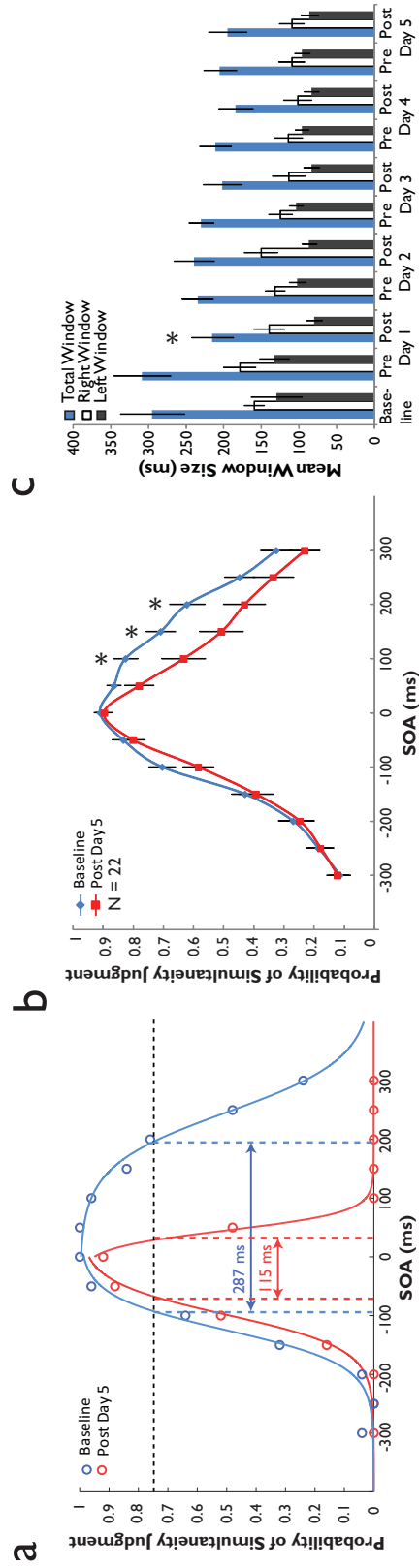


Figure 2.2 Training on a 2-AFC simultaneity judgment task narrows the temporal window of multisensory integration. a. Two sigmoid curves were fit to each individual's single-assessment data to derive an estimation of the temporal binding window (criterion=75%). In this individual, the size of the window narrows from 287 ms to 115 ms after 5 hours of feedback training. **b.** After training on the 2-AFC paradigm, 22 participants showed significant decreases in their probability of judging non-simultaneous audiovisual pairs as simultaneous. **c.** Average window sizes (total, left, and right) across the five days of training. Note that window size dropped significantly after the first hour of training, after which it remained stable. Error bars indicate 1 SEM. *, $p < 0.05$.

individual's initial assessment the mean span for the multisensory temporal window at this criterion was 287 ms (blue points and curve).

Perceptual training on a 2-AFC task results in a significant narrowing of the multisensory temporal binding window

Immediately following the training period there was a dramatic shift in judgments of simultaneity (**Fig. 2.2**). In the individual shown in **Figure 2.2a**, this was manifested as a decrease in window size from 287 ms at baseline assessment to 115 ms at the post-training assessment on day 5. Comparisons of group pre- and post-training simultaneity judgments at each SOA also reveal a strong effect (**Fig. 2.2b**). The largest training-related effects were seen on the right side of the distributions, corresponding to those conditions in which the visual stimulus preceded the auditory stimulus: for all stimulus conditions, a repeated measures ANOVA with within-subject factors SOA and Pre-Post Status resulted in a significant interaction ($F_{12, 238} = 10.11, p = 0.005$). Post-hoc paired-samples t-tests revealed significant decreases in mean probability judgment at the 100 ms (from 0.826 to 0.633, $p = 0.025$), 150 ms (from 0.709 to 0.507, $p = 0.016$), and 200 ms (from 0.622 to 0.431, $p = 0.020$) SOA conditions after correction for multiple comparisons. Hence, the training effect appears to be driven largely by significant decreases in the probability of simultaneity judgment following training at the objectively non-simultaneous conditions.

To examine the time course of the training-induced changes, we examined the simultaneity distributions at each of the eleven assessments completed throughout the course of training. Quite surprisingly, the effect is evident after a single day of training and is also equivalent in magnitude to that seen after 5 days of training (**Fig. 2.2c**).

Statistical analysis by repeated measures ANOVA revealed a significant main effect of assessment number ($n = 19$, $F_{10, 179} = 3.459$, $p < 0.001$) and post-hoc paired-samples t -tests with correction for multiple comparisons indicated a significant reduction in total window size from baseline assessment (mean of 294.59 ms) to post-training assessment day 1 (mean of 215.02 ms, $p = 0.045$, corrected). Window size did not differ significantly from post-training day 1 assessment onward (by repeated-measures ANOVA, $F_{8, 143} = 1.566$, $p = 0.140$), although means decreased from 215.02 ms to 194.87 ms. Interestingly, changes in window size seem to be wholly attributable to decreases in the right side of the temporal window. Thus, whereas the left side of the distribution did not change significantly over the course of training ($F_{10,179} = 1.637$, $p = 0.099$), the right side showed strong training-related changes (from 159.57 ms at baseline to 109.56 ms at post-training day 5 assessment, $F_{10,179} = 4.360$, $p = 1.77 \times 10^{-5}$).

Changes in the multisensory temporal window are not seen following passive exposure to the identical stimuli

As with the training group, the data generated by the 14 exposure control participants during each assessment were fitted with two sigmoid curves and window sizes were derived. **Figure 2.3a** shows data from a typical participant. Note that, in striking contrast to **Figure 2.2a**, the size of this individual's window appears to have increased after exposure (from 383 ms at baseline to 443 ms at post-exposure day 5 assessment). This change parallels similar effects on the group level (**Fig 2.3b**): comparison by repeated-measures ANOVA revealed a significant interaction between SOA and pre/post exposure status ($F_{12,142} = 7.793$, $p = 0.015$) and while uncorrected post-

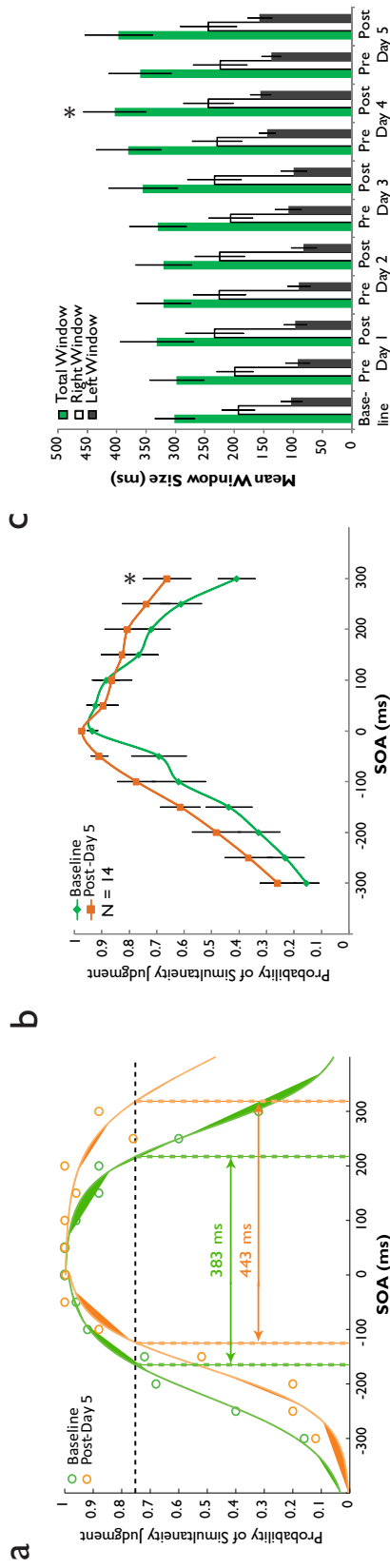


Figure 2.3. Changes in the multisensory temporal window are not seen following passive exposure to the identical stimuli.
a. As in the training participants, individual single-assessment data were fit with two sigmoid curves to derive an estimation of the temporal binding window in exposure controls. **b.** Control participants ($n=14$) exposed to the same stimulus pairs as training participants showed a significant increase in probability of simultaneity judgment after exposure. **c.** This was manifested as a significant increase in window size over the course of a week of exposure. Error bars indicate 1 SEM. *, $p < 0.05$

hoc t-tests indicate a significant *increase* in the probability of simultaneity judgment after exposure on the -50 ms, 0 ms, 200 ms, 250 ms, and 300 ms conditions (all $p < 0.05$), only the 300 ms condition shows a change after correction for multiple comparisons (from a mean of 0.409 at baseline to 0.663 at post-training day 5, $p = 0.002$). These results are upheld in analysis of group window size progression (**Fig. 2.3c**): one-way repeated-measures ANOVA indicated a main effect of assessment number ($n = 11$, $F_{10,99} = 2.212$, $p = 0.025$), with a significant increase in mean window size (from 301.21 ms at baseline to 403.33 ms at post-exposure day 5) first appearing at post-exposure assessment on day 4 ($p = 0.044$, corrected). However, this increase did not remain significant at pre- ($p = 0.660$, corrected) and post-exposure ($p = 0.069$, corrected) day 5 assessments. This difference in total window size appeared to be driven by an increase on the left side of the distribution ($F_{10,99} = 2.518$, $p = 0.011$). In contrast, analysis of the right side of the distribution indicated no effect ($F_{10,99} = 0.771$, $p = 0.656$).

Training-induced changes in the multisensory temporal binding window are stable for at least one week

Follow-up assessments were conducted on a subset of the participants in the training group ($n=16$) one week after the completion of training. Participants underwent no additional assessments or training during this week. Analysis of group-level probability of simultaneity judgment at each SOA tested (**Fig. 2.4a**) revealed a significant interaction between SOA and pre/post training status (repeated measures ANOVA, $F_{12,166} = 6.394$, $p = 0.023$) and post-hoc t-tests corrected for multiple comparisons revealed a number of significant decreases on the right side of the distribution (100 ms,

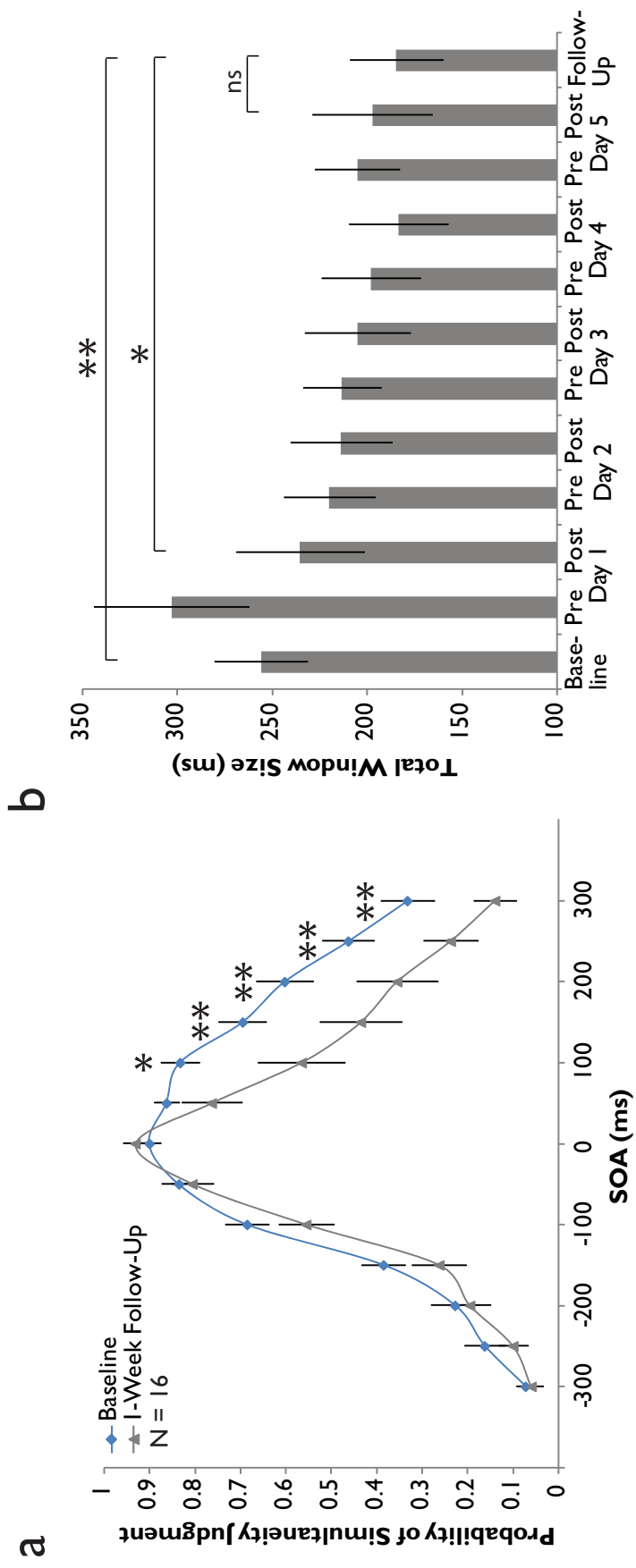


Figure 2.4. Training-induced changes in the multisensory temporal window are stable. a. Grand averaged data from 14 of the original 22 training subjects upon 1-week follow-up assessment compared to baseline. Significant decreases in probability of simultaneity judgment are seen exclusively on the right side of the distribution. **b.** Assessment across the 14 subjects shows a stable narrowing of window size from baseline. Error bars indicate 1 SEM. *, $p < 0.05$, **, $p < 0.01$.

from 0.833 to 0.565, $p = 0.016$; 150 ms, from 0.695 to 0.435, $p = 0.009$; 200 ms, from 0.603 to 0.355, $p = 0.010$; 250 ms, from 0.463 to 0.238, $p = 0.004$; 300 ms, from 0.333 to 0.140, $p = 0.004$). Analysis of window size change corroborates these results (**Fig. 2.4b**), indicating an overall effect of assessment number in follow-up participants ($n = 14$; repeated-measures ANOVA, $F_{10,129} = 3.873$, $p = 6.62 \times 10^{-5}$). Further analysis demonstrated 1-week follow-up window size (mean of 184.69 ms) to be significantly smaller than that at baseline assessment (mean of 255.86 ms, $p = 0.004$) as well as at post-training day 1 assessment (mean of 235.24 ms, $p = 0.039$), but not significantly different from post-training assessment on day 5 (mean of 197.15 ms, $p = 0.608$). Thus, while training-induced narrowing remained unchanged one week after training cessation, there was evidence of continued narrowing after the initial post-training day 1 drop.

2-IFC

Perceptual training on a 2-IFC simultaneity judgment task results in a significant narrowing in the size of the multisensory temporal binding window

While the 2-AFC results indicate a substantial, rapid, and long-lasting alteration in the size of the multisensory temporal binding window after perceptual training, it is possible that the effects seen may be driven, at least in part, by changes in cognitive biases (i.e., criterion shifts) associated with the two alternative design rather than by changes in sensory perceptual processes. To address this possibility, a cohort of 20 participants was recruited to take part in a two-interval forced choice (2-IFC) task wherein they were instructed to determine which of two sequential presentations of audiovisual pairs were simultaneous. This experimental structure does not require the

setting of a cognitive criterion for simultaneity and thus is more likely to reveal true improvements in discrimination abilities following perceptual training on the same task. It also carries with it the additional benefit of having a constant 1:1 ratio of simultaneous-to-non-simultaneous presentations, eliminating the need to alter this ratio for the training portion of the study.

As was done for the 2-AFC task results, individual data for each of the 20 subjects' assessments were fit with two sigmoid curves. Similarly to the procedure used to determine window size in the 2-AFC task, the value of the temporal binding window was set as the breadth of the distribution (in ms) at which individual participants performed at a criterion defined as halfway between their lowest accuracy point at baseline and 1 (the mean criterion level was 72.3%). **Figure 2.5a** illustrates the results of this process in one individual. Note that this individual's window size narrows from 349 ms to 182 ms following training. Group grand mean SOA analysis (**Fig. 2.5b**) revealed no overall effect of training at individual SOAs (repeated measures ANOVA, $F_{11,196}=0.792$, $p = 0.385$), likely the result of high inter-subject variability and the presence of individuals who fail to "learn" after training; see **Fig. 2.8**). However, analysis of window size as a function of training day (**Fig. 2.5c**) revealed a highly significant main effect of assessment number ($n = 17$; $F_{10,159}=4.503$, $p = 1.31 \times 10^{-5}$), with the first significant drop occurring between the post-training assessment on day 1 (mean of 275.79 ms) and the pre-training assessment on day 2 (mean of 173.67 ms, $p = 0.002$, corrected). Window size measures from this time period forward did not differ significantly ($F_{7,111}=2.067$, $p = 0.052$), and all remained significantly lower than baseline

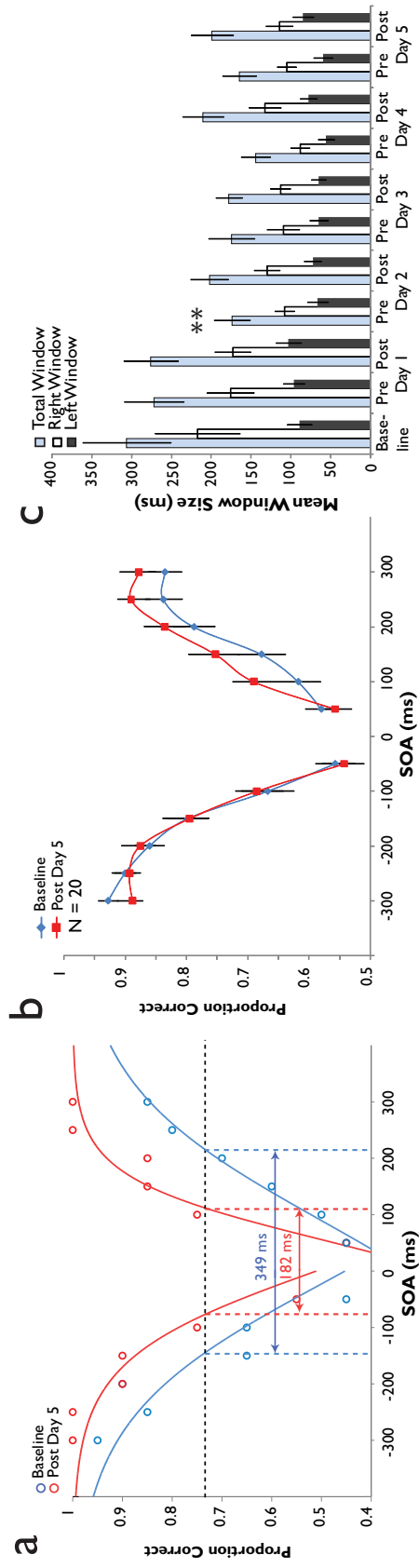


Figure 2.5. Perceptual Training on a 2-IFC simultaneity judgment task significantly alters the size of the multisensory temporal binding window. *a.* Individual single-assessment data fit with two sigmoid curves, from which a measure of the multisensory temporal binding window was derived (criterion: 75%). *b.* After 5 hours of training on the 2-IFC paradigm, 20 participants exhibited a narrowing of their temporal binding windows. *c.* Average window sizes (total, left, and right) across the five days of training. Note that window size dropped significantly between the first and second days of training, after which it remained stable. Error bars indicate 1 SEM. *, $p < 0.05$, **, $p < 0.01$.

(mean of 305.82 ms). In a striking similarity to the data derived from the 2-AFC portion of the study, the changes yielded by the 2-IFC training task seemed to be driven almost completely by shifts in the right side of the multisensory temporal distributions: although repeated-measures ANOVA indicated significant variation in left window size among the different assessments ($F_{10,159}=3.442$, $p<0.001$), no window size measurement proved to be statistically significant from baseline after correction for multiple comparisons. In contrast, window sizes on the right side varied significantly across assessments ($F_{10,159}=3.450$, $p < 0.001$) and, like the pattern in total window size change, first varies significantly from baseline (mean of 216.97 ms) at pre-training assessment on day 2 (mean of 107.68 ms, $p = 0.002$, corrected).

Changes induced by perceptual training on a 2-IFC simultaneity judgment task are stable for at least one week

One week after completion of their training on the 2-IFC task, 9 participants returned to complete a final assessment, the results of which are depicted in **Figure 2.6**. Analysis of grand mean accuracy as a function of SOA (**Fig. 2.6a**) revealed a significant interaction between SOA and pre/post training status ($F_{11,75} = 13.131$, $p = 0.007$), and post-hoc t-tests showed statistically significant increases in mean accuracy at 150 ms (from 0.681 to 0.881, $p = 0.010$, corrected), 200 ms (from 0.794 to 0.900, $p = 0.047$, corrected), and 300 ms (from 0.794 to 0.950, $p = 0.018$, corrected) SOA conditions. Once again, these effects were also evident on individual window size analysis (**Fig. 2.6b**), which indicated that there was significant variation among window sizes across all

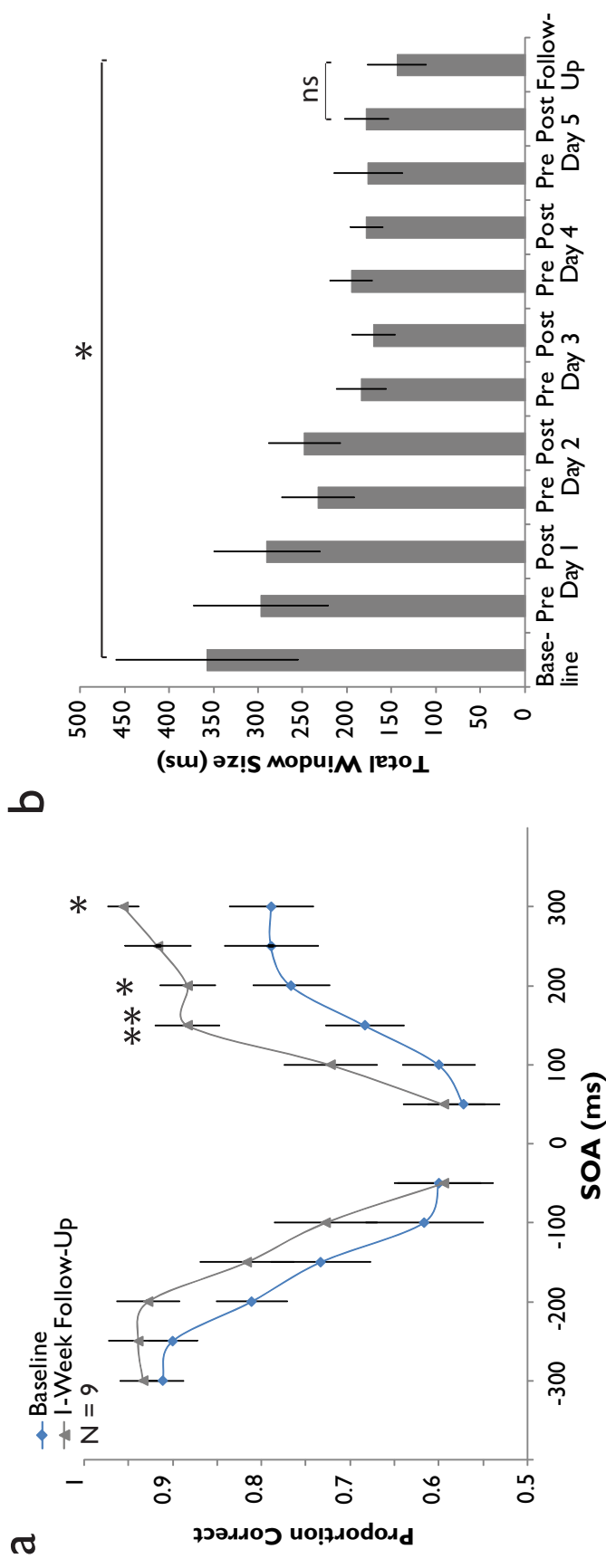


Figure 2.6. Changes induced by perceptual training on a 2-IFC simultaneity judgment task are stable over time. a. Averaged data from twelve 2-IFC subjects upon 1-week follow-up assessment. b. Average window size at follow-up is smaller than that at baseline assessment but not significantly different from that at post-training day 5 assessment. Error bars indicate 1 SEM. *, $p < 0.05$, **, $p < 0.01$.

assessments (repeated-measures ANOVA, $n = 7$, $F_{10,59} = 2.29$, $p = 0.019$), and that window size on one-week follow-up assessment (144.22 ms) was still significantly smaller than at baseline (357.53, $p = 0.001$) but was not significantly different from post-training day 5 assessment (178.17 ms, $p = 0.582$).

The window narrowing produced by 2-AFC and 2-IFC training tasks are highly similar in both degree of narrowing and its temporal dynamics

Examination of window size change over the course of training for both the 2-AFC and 2-IFC tasks allows for comparison of the dynamics of changes brought about by training under the two paradigms. **Figure 2.7** highlights several similarities and differences between the two groups. Although training utilizing the 2-AFC task results in a window size that is significantly narrower than baseline earlier (post-training day 1) than training under the 2-IFC task (pre-training day 2), and although the mean window size for the 2-IFC group is often lower than that of the 2-AFC after baseline assessment, an ANOVA with between-subject factor Group and within-subjects factor Assessment Number indicated no main effect of group ($F_{1,19} = 2.673$, $p = 0.103$) and no interaction between group and assessment number ($F_{10,188} = 0.993$, $p = 0.449$). On the level of each individual assessment, 2-IFC window size (173.67 ms) only significantly smaller than 2-AFC window size (234.50 ms) at pre-training day 2 assessment ($p = 0.024$), and this difference does not withstand correction for multiple comparisons. Overall, the degree and time-course of narrowing are remarkably similar between the two paradigms.

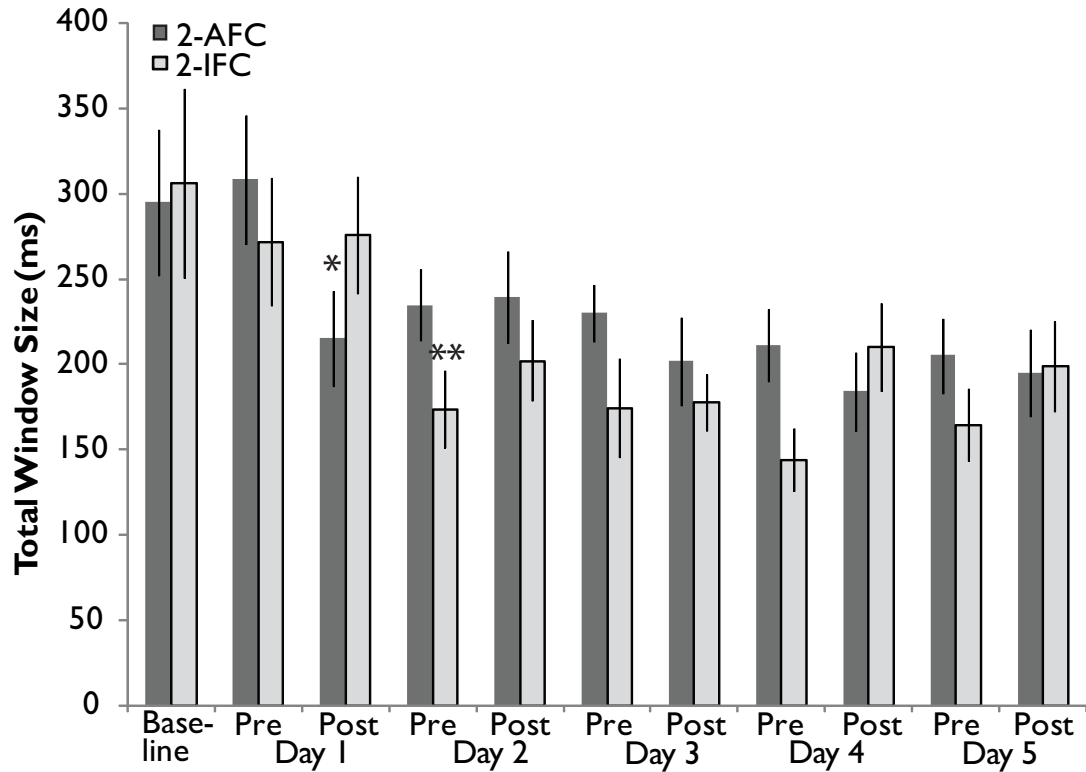


Figure 2.7. *The window narrowing produced by 2-AFC and 2-IFC training tasks are highly similar in dynamics and degree. Plotted are the average window sizes across the training week in 2-AFC (dark blue) and 2-IFC (light blue) subjects. Window size magnitude and degree of change do not significantly differ between the two groups, although 2-AFC reached significance from baseline one assessment earlier than those who underwent the 2-IFC task (asterisks). *, $p < 0.05$, **, $p < 0.01$.*

Large initial window size predicts success during training

It was noted upon analysis of the 2-IFC data that there appeared to be individuals whose mean window sizes decreased with training (dynamic participants) and those whose mean window sizes either remained the same or increased in size (static participants). 2-AFC dynamic participants' ($n = 13$) and static participants' ($n = 6$) window size progressions are plotted in **Figure 2.8a**. An ANOVA with between-subjects factor Group and within-subjects factor Assessment Number revealed a significant interaction ($F_{10,58} = 14.358$, $p < 0.001$), and the difference between groups at baseline assessment trended toward significance (dynamic participants, 344.01 ms, static participants, 187.40 ms, $p = 0.086$). Further analysis indicated that window size significantly decreased across the week's assessments in dynamic participants (repeated-measures ANOVA, $F_{10,119} = 4.125$, $p = 6.63 \times 10^{-5}$) but that no such change occurred in static participants' window size ($F_{10,49} = 0.737$, $p = 0.687$). **Figure 2.8b** highlights the differences in progression seen between 2-IFC dynamic participants ($n = 11$) and static participants ($n = 6$) over the course of the training week. Analysis of these differences by two-way ANOVA reveal a significant interaction between group and assessment number ($F_{10,58} = 2.318$, $p = 0.014$). Most importantly, the two groups differed significantly at baseline assessment (dynamic participants, 391.12 ms, static participants, 149.45 ms, $p = 0.032$), but did not differ significantly at any other individual assessment number. As was the case with 2-AFC dynamic participants, it was found that dynamic participants' window sizes decreased significantly over the course of the training week ($F_{10,99} = 5.656$, $p = 1.15 \times 10^{-6}$), but there were significant variations in window size in static participants over this period ($F_{10,49} = 3.604$, $p = 0.001$), driven by the sharp *increases* on post-training day 4 and post-training

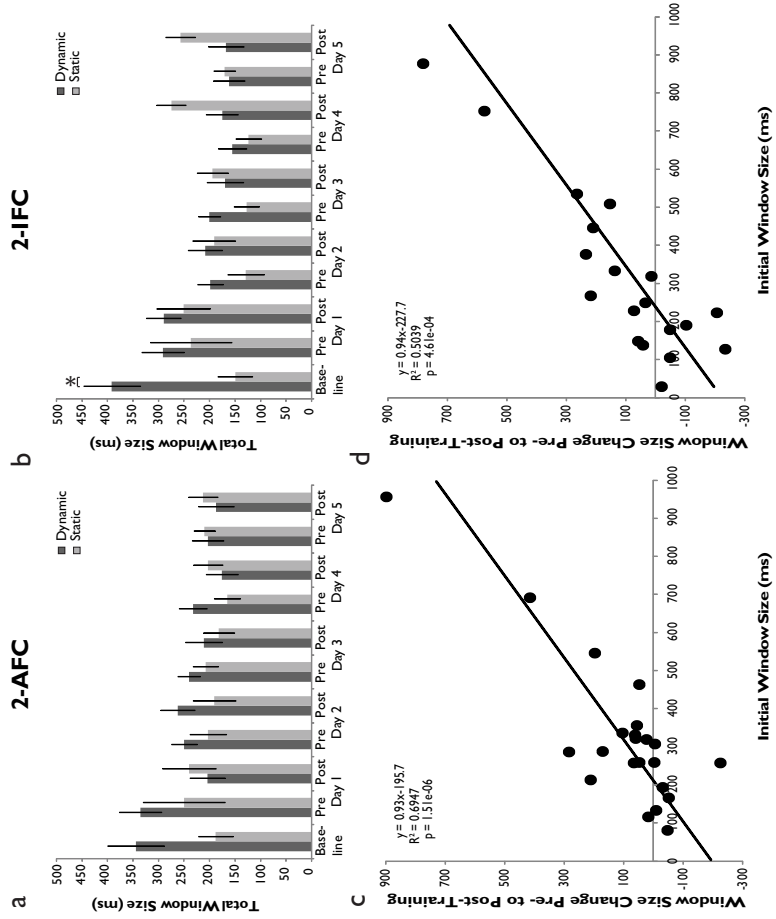


Figure 2.8. Large initial window size predicts success during training. **a.** Plotted mean total window size across all days of training in 2-AFC dynamic participants (whose window size at baseline is greater than at post-training day 5) and static participants (whose baseline window size is the same as or less than that at post-training day 5). Note that the difference in window size between these two groups is initially large and disappears over the course of the training week. **b.** Corresponding data from the 2-IFC group. Note both the large difference between groups at baseline as well as the sawtooth pattern that is absent in dynamic participants but present in static participants. **c.** Correlation plot of initial window size measurements versus window size change in 2-AFC training group. **d.** Corresponding data from the 2-IFC group. *, $p < 0.05$.

day 5 assessments. These increases form part of an overall pattern characterized by increased window sizes upon post-training assessment as compared to pre-training, leading to the appearance of a sawtooth pattern in the window sizes of static participants over the training week. Interestingly, this pattern does not appear in dynamic participants but is prominent in the 2-IFC group data (**Fig. 2.5c**), indicating that the latter effect may be wholly driven by the increases exhibited by the 2-IFC static participants.

Examination of individual subjects' initial window sizes and the window size changes exhibited by these individuals after training yielded significant correlations in both the 2-AFC ($R^2 = 0.695$, $p = 1.51 \times 10^{-6}$) and 2-IFC ($R^2 = 0.504$; $p = 4.61 \times 10^{-4}$) data sets. Even more striking, the lines of best fit for these data sets have very similar slopes (0.93, 2-AFC; 0.94, 2-IFC) and x-intercepts converging near 200 ms, the approximate size of static participants' initial windows. Together, these analyses indicate that it is possible to predict the direction and magnitude of window size change based upon initial window size.

Discussion

We have demonstrated that two multisensory perceptual training paradigms are capable of effecting significant, lasting changes in participants' judgments of the perceived simultaneity between visual and auditory events. Moreover, we have provided strong evidence that these effects are driven by a true change in perceptual discrimination abilities engendered by training, and are not a result of simple exposure to the repeated statistical regularities of the training stimuli.

Examination of the multisensory temporal window distributions prior to training on both tasks revealed a strong asymmetry, with a shoulder of increased probability of simultaneity judgment on the right half of the distribution (i.e., when the onset of the visual stimulus precedes the auditory stimulus). This asymmetry is consistent with other measures of the multisensory temporal window (McGurk and MacDonald, 1976; Dixon and Spitz, 1980; McGrath and Summerfield, 1985), and may be explained by the fact that the visual-leading conditions (unlike auditory-leading conditions) are ethologically valid and must be flexibly specified based upon the distance of the stimulus from the observer. This asymmetry is eliminated with training, most likely reflecting the symmetrical structure of the training tasks.

One of the most surprising effects of the perceptual training was its time course. In both the 2-AFC and 2-IFC tasks, significant effects emerged after a single day of training. Indeed, there is growing evidence that short-term exposure to asynchronous audiovisual pairs can drive temporal recalibration (Fujisaki et al., 2004; Vroomen et al., 2004; Navarra et al., 2005; Keetels and Vroomen, 2007; Navarra et al., 2007; Vatakis et al., 2007; Hanson et al., 2008; Keetels and Vroomen, 2008). These short-term effects have been shown to be transient, and as a consequence of these prior studies our expectation was that the effects of multisensory perceptual learning would not be retained long after the cessation of training. In contrast, the training effects in the current study showed a stability that extended at least a week after the cessation of training. Indeed, performance further improved during this week. In the 2-AFC task, this improvement on follow-up is seen not only as a decline in reports of simultaneity for non-simultaneous conditions, but also as a significant increase in the probability that participants will judge

the veridical simultaneous condition as simultaneous. Possible explanations for this improvement include recovery from fatigue associated with five consecutive days of training, a possibility that needs further investigation (see below). Another intriguing possibility is that long-term memory consolidation may play a role in strengthening the original training effects (see (McGaugh, 2000) for review). Indeed, several studies have demonstrated that perceptual learning effects may depend greatly upon sleep-mediated consolidation (Karni et al., 1994; Maquet, 2001; Fenn et al., 2003; Walker and Stickgold, 2004).

The results of the passive exposure experiment deserve particular note, given the surprising finding of an increase in window size over the course of the week of exposure to training stimuli. Because the switch from assessment to training/exposure in the 2-AFC task required an alteration in the ratio of simultaneous to non-simultaneous stimulus presentations from 1:1 to 6:1, the widening of the temporal binding window observed in exposure subjects may well represent an implicit learning phenomenon, wherein participants “learn” during exposure that an increased number of simultaneous presentations is occurring and subsequently bias their responses. This hypothesis is supported by similar data shown in **Figure 2.9**, derived from a small number of participants (n=5) who underwent the 2-AFC training paradigm without any explicit feedback; thus, these participants did not perform the exposure (oddball) task, but performed the original training task without the presence of feedback. These data show a similar (but not statistically significant) increase in window size over the course of the week. In addition, these results indicate that the training effects observed were not the result of a simple narrowing of the stimulus space (from 300 ms on each side of the

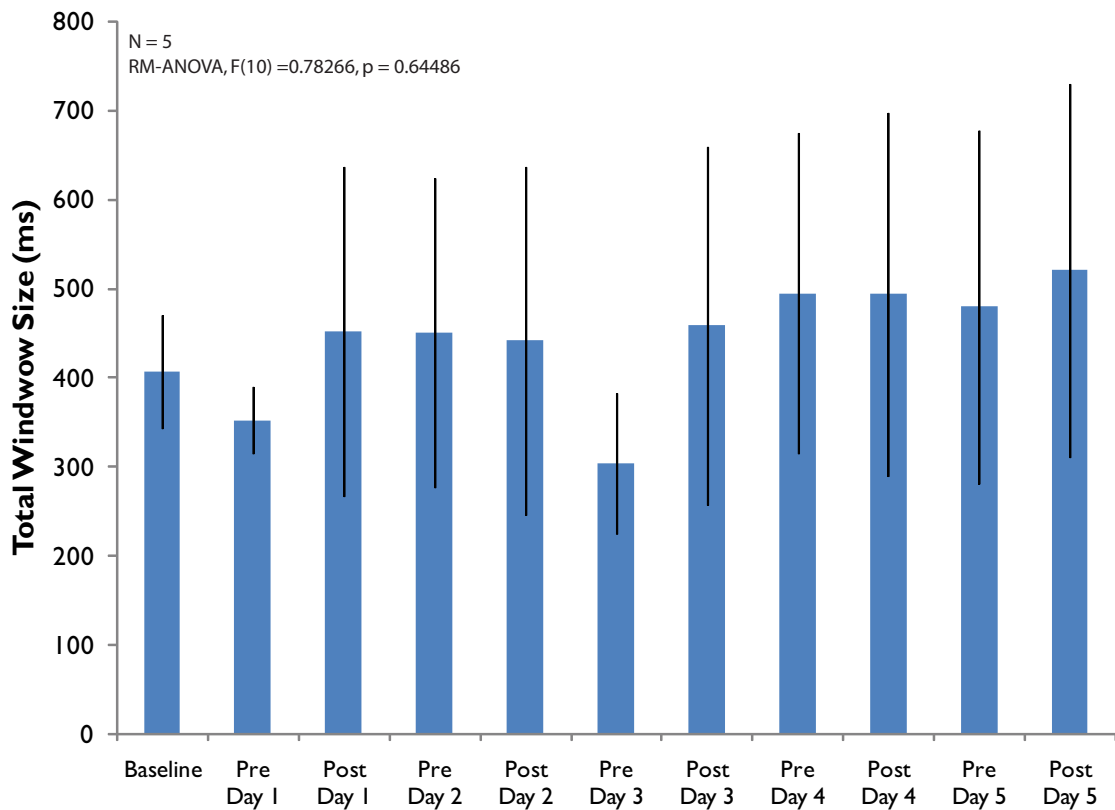


Figure 2.9. Training without feedback fails to narrow the multisensory temporal binding window. Plotted are data from 5 participants who underwent the same training protocol as the 2-AFC training group but in the absence of explicit feedback. Results indicate that window size remains stable in the absence of feedback.

distribution during assessment to 150 ms on each side during training) but were the true result of feedback training.

The importance of feedback for the observed narrowing of the multisensory temporal binding window fits well with what is known about the critical elements for engaging sensory plasticity. Seminal studies showed that significant reorganization could be driven in a bottom-up fashion by exposure to a constrained set of sensory stimuli early in development (Hubel et al., 1977; Simons and Land, 1987; Zhang et al., 2001; de Villers-Sidani et al., 2007; Han et al., 2007), and that passive exposure to these same stimuli became less likely to drive behavioral change and neural reorganization as an animal reached the end of a critical period of development (Hubel and Wiesel, 1963). Later studies revealed that these anatomical, behavioral and physiological changes induced in developing animals by passive exposure could indeed take place in adults via top-down perceptual learning, wherein stimuli are paired with either positive or negative reinforcement (Salazar et al., 2004; Blake et al., 2006; Polley et al., 2006). Thus, it appears that the pairing of an alteration in the sensory statistics with an instructive signal (i.e., feedback) is crucial for adult sensory reorganization, and this principle is supported by the data reported here.

Further work is needed to better characterize the effects of fatigue and inattention on the size of the multisensory temporal binding window. As highlighted in the data, the increases in window size on each post-training assessment in the 2-IFC task are driven exclusively by static participants whose temporal windows are small prior to any training. This fact points strongly to the idea that these increases may be the result of fatigue or inattention and is further supported by an analysis of response bias in these individuals

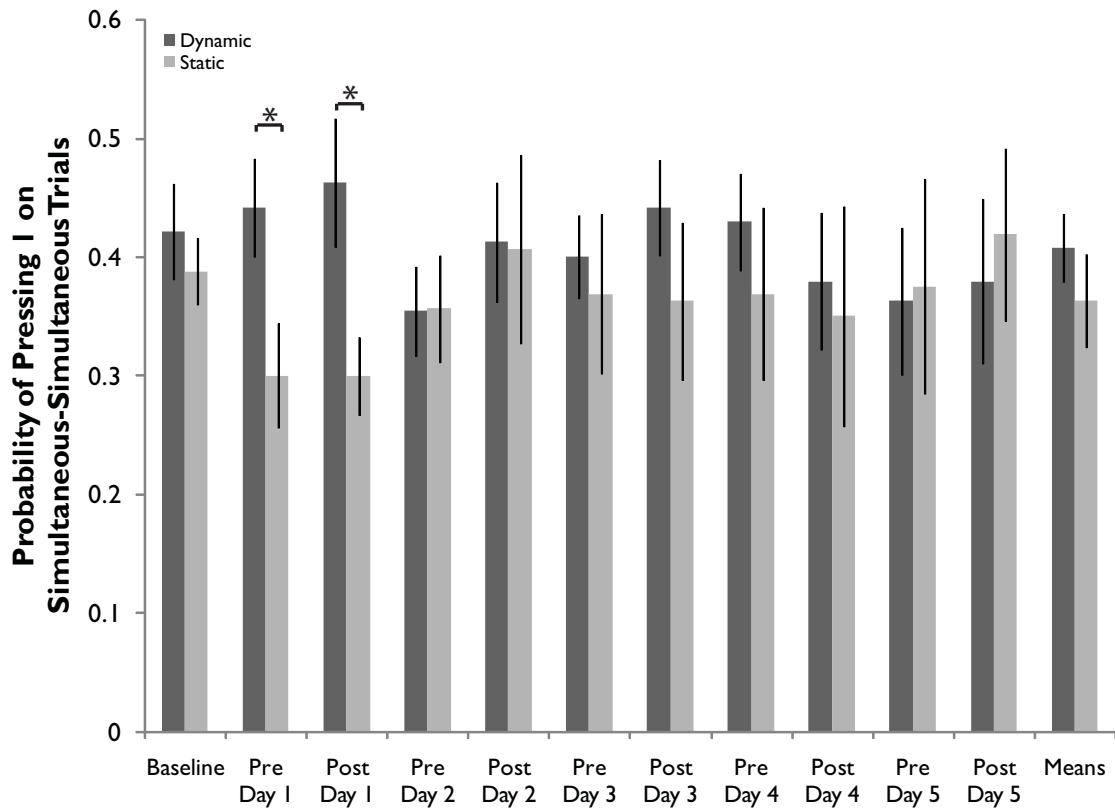


Figure 2.10. *Static participants on the 2-IFC task show increased response bias on some assessments. The probability that dynamic participants (dark gray) and static participants (light gray) will press “1” on simultaneous-simultaneous catch trials is plotted over the course of the training week. Although all participants show a significant bias toward pressing “2” on these trials, this bias is most pronounced in static participants at a point of high fatigue (the end of the first day of training). Note also that these points also correspond to assessments at which static participants show increases in window size (Fig. 8).*

(**Fig. 2.10**). These data were derived from responses in 2-IFC catch trials during which participants were presented with two simultaneous (instead of one simultaneous and one non-simultaneous) audiovisual events. Results show that, while all individuals appear to share a bias toward indicating that the simultaneous pair was present in the second interval during this condition ($p = 1.421 \times 10^{-4}$), static participants have a much more pronounced bias than dynamic participants during the pre- ($p = 0.035$) and post-training ($p = 0.037$) assessments on day 1, during which participants spent the most time in the lab and showed large increases in window size (**Fig. 2.8**).

Importantly, and despite the above considerations, several pieces of evidence indicate that the phenomena observed here reflect changes in sensory perceptual rather than cognitive systems. First is the fact that, by and large, the size of the multisensory temporal binding window at baseline assessment and the narrowing brought about by training are remarkably similar in spite of alterations in task structure, pointing to a construct that is driven largely by changes in (multi)sensory representations. It should also be noted that the mean window sizes and shapes observed at baseline assessment (~275-300 ms, skewed toward the positive side of the distribution) are very much in accord with those reported in prior work (Dixon and Spitz, 1980; Bushara et al., 2001; Zampini et al., 2003; Colonius and Diederich, 2004; Fujisaki et al., 2004; Zampini et al., 2005b; Zampini et al., 2005a), making the training effects reported here even more striking in that they alter both of these characteristics of the temporal binding window. Second, and most notably, trends across the span of the training week indicate that dynamic participants tend to exhibit larger temporal windows at baseline assessment than do static participants and that dynamic participants' windows narrow until they are at a

size that is comparable to those of static participants. Taken together, these results indicate that there may be a lower limit to the size of multisensory binding window in typically-developing adults. However, further studies must be done to rule out the possibility that this seeming lower limit is an artifact of the specifics of the tasks' training and reward structure.

A search for the neural bases of the multisensory temporal binding window described here has become an increasingly active area of inquiry of late, and can trace its origin to earlier studies that highlighted the importance of temporal factors in modulating multisensory integration at the level of the single-cell (Meredith et al., 1987b; Wallace et al., 1996; Stanford et al., 2005). This work has now been extended to the network level, where a number of recent studies point to the presence of a large, dynamic network of areas that include the insula, posterior parietal and superior temporal cortices as being critical in the perception of audiovisual simultaneity (Bushara et al., 2001; Calvert et al., 2001; Bushara et al., 2003; Noesselt et al., 2007; Noesselt et al., 2008). Most recently, interest has been focused on the potential role of neuronal oscillations in multisensory processing and temporal binding (Lakatos et al., 2007; Chandrasekaran and Ghazanfar, 2009). Taken together, this work points to cortex as the critical locus for perceptual plasticity (Schwartz et al., 2002; Pleger et al., 2003; Maertens and Pollmann, 2005).

Returning to the single cell, one readily envisioned mechanism to subserve the plastic changes in evidence here is a narrowing in the temporal tuning profile of multisensory neurons responsible for binding processes. In virtually all studies examining these tuning functions in individual multisensory neurons, the temporal windows within which significant multisensory interactions can be generated have been shown to be

surprisingly wide (i.e., several hundred milliseconds; see (Meredith et al., 1987b; Wallace et al., 1996; Stanford et al., 2005)). Although these tuning functions have been shown to be relatively static in adults even in the face of significant changes in sensory statistics (Polley et al., 2008), the coupling with a reinforcement-based signal may be sufficient to engender significant change. Indeed, physiological studies of adult plasticity within sensory systems have focused on basal forebrain cholinergic signals as the instructive cue (Huerta and Lisman, 1995; Hohmann and Berger-Sweeney, 1998; Kilgard, 2003).

Moving beyond the single cell, another plausible mechanism is a consolidation in the timing circuits that serve to perceptually anchor stimulus events from different modalities, with this consolidation serving to narrow the tolerance for the encoding of unity judgments. Whether the critical consolidation takes place in one of the nodes in the cortical network, or whether it is distributed awaits future study. A final potential mechanism could feature changes in oscillatory patterns within or across cortical domains that are integral in temporal binding. In addition to the role of cortex, increasing evidence indicates that subcortical structures are far more capable of plastic change than previously thought (Illing, 2001; de Boer and Thornton, 2008; Song et al., 2008; Tzounopoulos and Kraus, 2009). Hence, future neuroimaging studies will focus on both cortical and subcortical structures in order to elucidate the neural bases of the temporal plasticity evident in multisensory systems.

Overall, the results reported here indicate that training on a simultaneity judgment task is capable of eliciting meaningful, lasting changes in the size of individuals' multisensory temporal binding windows. This ability holds particular promise in

designing remediation strategies for disorders (i.e., dyslexia, autism, schizophrenia) in which altered multisensory temporal processing is a contributory factor.

Acknowledgments

This research was supported by the Vanderbilt Kennedy Center for Research on Human Development, the National Institute for Child Health and Development (HD050860), and the National Institute for Deafness and Communication Disorders (F30 DC009759). We would like to thank Drs. Calum Avison, Randolph Blake, Maureen Gannon, and Daniel Polley, as well as Leslie Dowell, Matthew Fister, Matthew Hevey, Haleh Kadivar and Juliane Krueger for their technical, conceptual and editorial assistance.

References

Blake DT, Heiser MA, Caywood M, Merzenich MM (2006) Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. *Neuron* 52:371-381.

Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300-304.

Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M (2003) Neural correlates of cross-modal binding. *Nat Neurosci* 6:190-195.

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14:427-438.

Chandrasekaran C, Ghazanfar AA (2009) Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. In, p 773: Am Physiological Soc.

Colonius H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci* 16:1000-1009.

de Boer J, Thornton AR (2008) Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J Neurosci* 28:4929-4937.

de Gelder B, Vroomen J, Annen L, Masthof E, Hodiament P (2003) Audio-visual integration in schizophrenia. *Schizophr Res* 59:211-218.

de Gelder B, Vroomen J, de Jong SJ, Masthoff ED, Trompenaars FJ, Hodiament P (2005) Multisensory integration of emotional faces and voices in schizophrenics. *Schizophr Res* 72:195-203.

de Villers-Sidani E, Chang EF, Bao S, Merzenich MM (2007) Critical Period Window for Spectral Tuning Defined in the Primary Auditory Cortex (A1) in the Rat. In: *J. Neurosci*, pp 180-189.

Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9:719-721.

Fenn KM, Nusbaum HC, Margoliash D (2003) Consolidation during sleep of perceptual learning of spoken language. In, pp 614-616: Nature Publishing Group.

Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7:773-778.

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT (2005) Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res* 166:474-480.

Han YK, Kover H, Insanally MN, Semerdjian JH, Bao S (2007) Early experience impairs perceptual discrimination. *Nat Neurosci* 10:1191-1197.

- Hanson JV, Heron J, Whitaker D (2008) Recalibration of perceived time across sensory modalities. *Exp Brain Res* 185:347-352.
- Hohmann CF, Berger-Sweeney J (1998) Cholinergic regulation of cortical development and plasticity. *New twists to an old story*. In, p 401.
- Hubel DH, Wiesel TN (1963) Receptive Fields of Cells in Striate Cortex of Very Young, Visually Inexperienced Kittens. *J Neurophysiol* 26:994-1002.
- Hubel DH, Wiesel TN, LeVay S (1977) Plasticity of ocular dominance columns in monkey striate cortex. *Philos Trans R Soc Lond B Biol Sci* 278:377-409.
- Huerta PT, Lisman JE (1995) Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. In, pp 1053-1064: Cambridge, Mass.: Cell Press, c1988-.
- Illing RB (2001) Activity-dependent plasticity in the adult auditory brainstem. *Audiol Neurootol* 6:319-345.
- Karni A, Tanne D, Rubenstein BS, Askenasy JJ, Sagi D (1994) Dependence on REM sleep of overnight improvement of a perceptual skill. In, pp 679-682.
- Keetels M, Vroomen J (2007) No effect of auditory-visual spatial disparity on temporal recalibration. *Exp Brain Res* 182:559-565.
- Keetels M, Vroomen J (2008) Temporal recalibration to tactile-visual asynchronous stimuli. *Neurosci Lett* 430:130-134.
- Kilgard M (2003) Cholinergic modulation of skill learning and plasticity. In, pp 678-680: Elsevier.
- Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron* 53:279-292.
- Lewkowicz DJ (1996) Perception of auditory-visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform* 22:1094-1106.

- Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci U S A* 103:6771-6774.
- Lewkowicz DJ, Sowinski R, Place S (2008) The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. *Brain Res*.
- Maertens M, Pollmann S (2005) fMRI reveals a common neural substrate of illusory and real contours in V1 after perceptual learning. In, pp 1553-1564: MIT Press.
- Maquet P (2001) The role of sleep in learning and memory. In, pp 1048-1052.
- McGaugh JL (2000) Memory--a century of consolidation. *Science* 287:248-251.
- McGrath M, Summerfield Q (1985) Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J Acoust Soc Am* 77:678-685.
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746-748.
- Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365:350-354.
- Meredith MA, Nemitz JW, Stein BE (1987a) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215-3229.
- Meredith MA, Nemitz JW, Stein BE (1987b) Determinants of multisensory integration in superior colliculus neurons. 1 Temporal factors. *J Neuroscience* 7:3215-3229.
- Navarra J, Soto-Faraco S, Spence C (2007) Adaptation to audiotactile asynchrony. *Neurosci Lett* 413:72-76.
- Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Res Cogn Brain Res* 25:499-507.
- Noesselt T, Bonath B, Boehler CN, Schoenfeld MA, Heinze HJ (2008) On perceived synchrony—neural dynamics of audiovisual illusions and suppressions. In, pp 132-141: Elsevier.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. In, p 11431: Soc Neuroscience.

Pleger B, Foerster AF, Ragert P, Dinse HR, Schwenkreis P, Malin JP, Nicolas V, Tegenthoff M (2003) Functional imaging of perceptual learning in human primary and secondary somatosensory cortex. In, pp 643-653: Elsevier.

Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26:4970-4982.

Polley DB, Hillock AR, Spankovich C, Popescu MV, Royal DW, Wallace MT (2008) Development and plasticity of intra-and intersensory information processing. In, p 780.

Salazar RF, Kayser C, Konig P (2004) Effects of training on neuronal activity and interactions in primary and higher visual cortices in the alert cat. *J Neurosci* 24:1627-1636.

Schwartz S, Maquet P, Frith C (2002) Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. In, pp 17137-17142: National Acad Sciences.

Simons DJ, Land PW (1987) Early experience of tactile stimulation influences organization of somatic sensory cortex. *Nature* 326:694-697.

Song JH, Skoe E, Wong PC, Kraus N (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 20:1892-1902.

Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. In, pp 6499-6508: Soc Neuroscience.

Teder-Salejarvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. *J Cogn Neurosci* 17:1396-1409.

Tzounopoulos T, Kraus N (2009) Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62:463-469.

- Vatakis A, Navarra J, Soto-Faraco S, Spence C (2007) Temporal recalibration during asynchronous audiovisual speech perception. *Exp Brain Res* 181:173-181.
- Virsu V, Lahti-Nuuttila P, Laasonen M (2003) Crossmodal temporal processing acuity impairment aggravates with age in developmental dyslexia. *Neurosci Lett* 336:151-154.
- Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res* 22:32-35.
- Walker MP, Stickgold R (2004) Sleep-dependent learning and memory consolidation. In, pp 121-133: Elsevier.
- Wallace MT, Stein BE (1997) Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci* 17:2429-2444.
- Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91:484-488.
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. In, pp 1246-1266: Am Physiological Soc.
- Wallace MT, McHaffie JG, Stein BE (1997) Visual response properties and visuotopic representation in the newborn monkey superior colliculus. *J Neurophysiol* 78:2732-2741.
- Zampini M, Shore DI, Spence C (2003) Audiovisual temporal order judgments. *Exp Brain Res* 152:198-210.
- Zampini M, Shore DI, Spence C (2005a) Audiovisual prior entry. *Neurosci Lett* 381:217-222.
- Zampini M, Guest S, Shore DI, Spence C (2005b) Audio-visual simultaneity judgments. *Percept Psychophys* 67:531-544.
- Zhang LI, Bao S, Merzenich MM (2001) Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nat Neurosci* 4:1123-1130.

CHAPTER III

NEURAL CORRELATES OF MULTISENSORY PERCEPTUAL LEARNING[†]

Abstract

The brain's ability to bind incoming auditory and visual stimuli depends critically on the temporal structure of this information. Specifically, there exists a temporal window of audiovisual integration within which stimuli are highly likely to be perceived as part of the same environmental event. Several studies have described the temporal bounds of this window, but few have investigated its malleability. Recently, our laboratory has demonstrated that a perceptual training paradigm is capable of eliciting a 40% narrowing in the width of this window that is stable for at least one week after cessation of training. In the current study we sought to reveal the neural substrates of these changes. Eleven subjects completed an audiovisual simultaneity judgment training paradigm, immediately before and after which they performed the same task during an event-related 3T fMRI session. The posterior superior temporal sulcus (pSTS) and areas of auditory and visual cortex exhibited robust BOLD decreases following training, and resting state and effective connectivity analyses revealed significant increases in coupling between these cortices after training. These results represent the first evidence of the neural correlates underlying plastic change in adult multisensory networks that likely represent the substrate for a multisensory temporal binding window.

[†] The contents of this chapter are included in a manuscript submitted to *Nature Neuroscience* and thus follow that journal's format.

Introduction

We live in a world rich in sensory information. As such, the degree to which we are able to parse and combine this information effectively determines our ability to perform many tasks, and continually shapes our perceptions of the world. To solve this important and computationally expensive set of problems the brain takes advantage of certain statistical regularities of stimuli within the physical world—including spatial and temporal congruity—in determining which information from the different senses should be perceptually bound. A number of studies have established that visual and auditory events emanating from the same location in space are likely to be perceptually bound (Meredith and Stein, 1986; Wallace et al., 1992; Teder-Salejarvi et al., 2005; Innes-Brown and Crewther, 2009). The same may be said in the temporal domain, but because the propagation speeds of the energies carrying visual and auditory information differ, some flexibility must be afforded in this determination. Thus, in describing whether or not asynchronously occurring auditory and visual events are likely to be perceptually unified, the concept of a temporal window of multisensory binding has become a useful construct. Within this time interval (generally in the range of 250 – 300 ms) auditory and visual events are likely to be perceptually bound and produce neurophysiological, behavioral, and perceptual changes (Meredith et al., 1987; Sekuler et al., 1997; Shams et al., 2002; Colonius and Diederich, 2004).

Recently, studies have begun to focus upon elucidating the neural correlates of the multisensory temporal binding window, and human neuroimaging studies have identified a network of brain regions that appear to be important in the perception of audiovisual temporal relations. In previous PET and fMRI studies, canonical regions of multisensory

convergence such as superior colliculus (Bushara et al., 2001; Calvert et al., 2001), posterior superior temporal sulcus (pSTS) (Calvert et al., 2001), inferior parietal lobule (Bushara et al., 2001), and insula (Bushara et al., 2001; Calvert et al., 2001) have been shown to be preferentially active during the presentation of synchronous auditory and visual stimuli (as compared to asynchronous presentation). More recent work has extended this network to include domains within visual and auditory cortices (Bischoff et al., 2007; Dhamala et al., 2007; Noesselt et al., 2007).

Although these imaging studies have identified the important nodes involved in multisensory temporal perceptual processes, their ability to clarify the functional roles these nodes may play in processing multisensory temporal information is limited. More functionally-oriented studies would strive to manipulate these networks in ways that capture the dynamics of the temporal encoding processes. Recent work has provided just such a tool by demonstrating that training on an audiovisual simultaneity judgment task is capable of eliciting a robust, rapid and stable narrowing of the multisensory temporal binding window (Powers et al., 2009). The current study takes advantage of this plasticity to uncover the dynamic interactions among the nodes of the network that may underlie perception of audiovisual simultaneity.

Results

Perceptual training narrows the multisensory temporal binding window

Eleven subjects participated in the psychophysical and neuroimaging portions of the study, which took place over two days. The study consisted of initial psychophysical and neuroimaging assessments, followed by a training phase and then final psychophysical and imaging assessments. In the assessments, participants engaged in a two-interval forced choice (2-IFC) audiovisual simultaneity judgment task (**Fig. 3.1a, b**) wherein they reported which of two presentations of an audiovisual stimulus pair occurred simultaneously (i.e., at a stimulus onset asynchrony [SOA] of 0 ms). Within the other interval the stimuli were always asynchronously presented, at SOAs ranging from -300 ms (auditory leading visual) to +300 ms (visual leading auditory) in 50 ms increments. Following the initial assessment, participants were then trained on the same task. The training portion used identical stimuli but provided feedback on each trial as to the correctness of the participant's response. Neuroimaging assessments consisted of high-resolution T1 structural scans, resting state functional connectivity, blocked multisensory effective connectivity, and event-related fMRI scans wherein participants performed the 2-IFC task. For practical reasons, fMRI data was only collected for SOAs of 0 ms (objective simultaneity), 300 ms (largest asynchrony) and an intermediate SOA determined for each participant and which represented a value at or near the border of their individualized multisensory temporal binding window (labeled as Right Window Size, RWS).

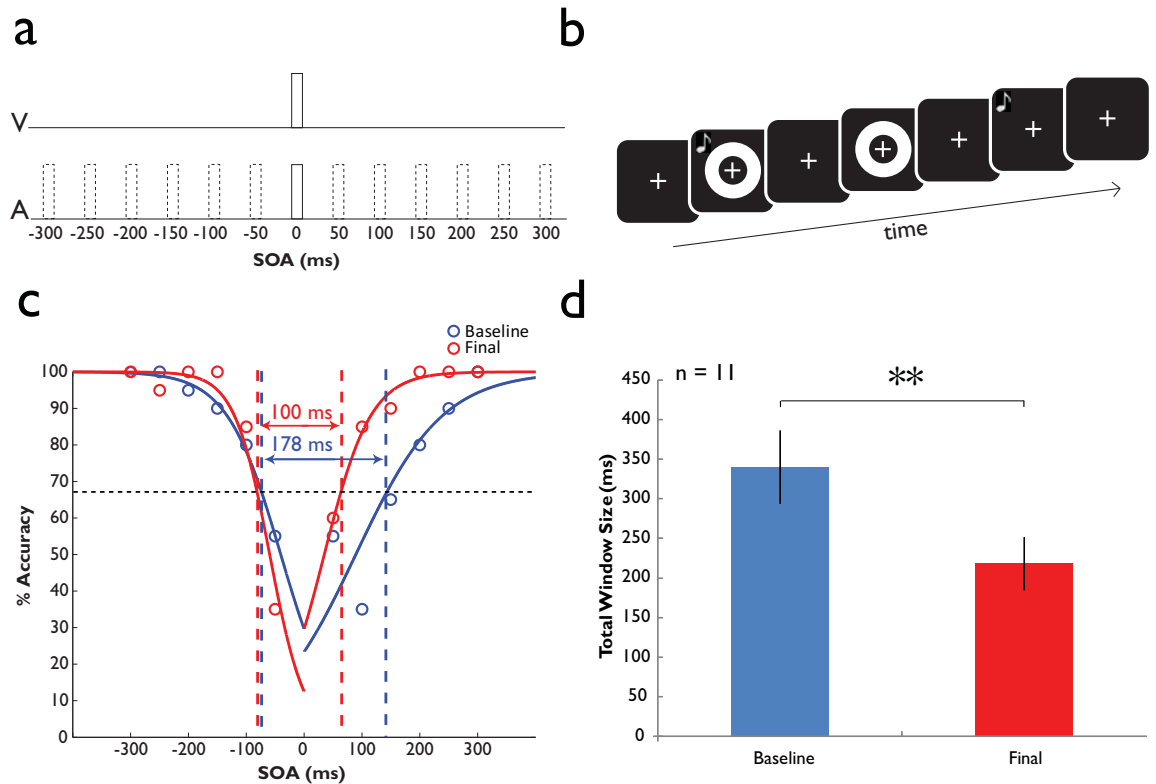


Figure 3.1. Experimental procedure and behavioral results. *a.* Stimulus presentation protocol for a forced choice trial. Here, one stimulus pair is always simultaneously presented (stimulus onset asynchrony, SOA = 0), and one is separated by some SOA ranging from -300 ms (auditory leading visual) to +300 ms (visual leading auditory), by 50 ms increments. *b.* Physical characteristics and temporal structure of the stimuli used. Depicted is a simultaneous-first two-interval forced choice trial. *c.* Data from one participant at baseline and final assessments, plotted as mean accuracy as a function of stimulus onset asynchrony (SOA). These data are fitted with two sigmoid curves to model the left and right sides of the temporal binding window, and the size of this window is taken to be the breadth of this distribution at half the distance between the minimum data point at baseline and 1 (full width at half height). By this measure, this individual's temporal window size narrows from 178 ms at baseline to 100 ms after training. *d.* Mean total window size from all 11 participants. Mean window size decreases significantly from baseline (340 ms) to final assessment (219 ms). Error bars indicate one S.E.M.; **, $p < 0.01$.

In accordance with our prior psychophysical study (Powers et al., 2009), analysis of data from the participants who underwent training on the 2-IFC task revealed a marked narrowing of their multisensory temporal binding window. **Figure 3.1c** features data from one individual at baseline and upon final assessment. Note the significant improvements in performance following training, and the decline in the width of the temporal window (see methods for details on how these windows were delimited) from 178 ms at baseline to 100 ms at final assessment. **Figure 3.1d** depicts the significant change in the group window size following training. Whereas baseline window size was 340 ms, following training this value declined to 219 ms (paired samples *t-test*, $p = 0.00759$, corrected). Both the mean window size at baseline and the degree of narrowing exhibited by participants following training (36% decrease from baseline) are very similar to what has been previously described (Powers et al., 2009) and illustrates the efficacy of the perceptual training regimen in altering audiovisual simultaneity perception.

Multisensory timing-dependent networks change with perceptual training

Important in the determination of which brain networks may underlie the improvements in multisensory temporal perception is the identification of brain areas that respond differentially to synchronous and asynchronous presentations of audiovisual stimulus pairs. To create this contrast, two event-related runs from the pre-training session and two from the post-training session were isolated and used as a localizer, contrasting activity related to synchronous presentation with that related to asynchronous presentation. The localizer identified a network of areas known to be important for

audiovisual synchrony perception, including the multisensory areas pSTS (Calvert et al., 2000; Calvert et al., 2001; Dhamala et al., 2007; Noesselt et al., 2007; Stevenson et al., 2010); insula (Calvert et al., 2001); posterior parietal cortex (Dhamala et al., 2007), and lateral occipital cortex (Dhamala et al., 2007), as well as visual (Noesselt et al., 2007; Stevenson et al., 2010) and auditory areas (Noesselt et al., 2007; Stevenson et al., 2010). Additionally, regions of the superior cerebellum closely associated with the nucleus interpositus, known from lesion and functional imaging studies to be important in sub-second timing tasks (Perrett et al., 1993; Bao et al., 2002; Kotani et al., 2003), were also identified.

After these initial analyses succeeded in identifying a plausible network of cortical areas that are differentially active to synchronous versus asynchronous audiovisual pair presentation, a second analysis sought to determine if activity in these areas changed after training. On the whole-brain level, only two contrasts yielded significant clusters of activation that overlapped with the localizer. These two contrasts were at objective simultaneity (Pre SOA 0 > Post SOA 0; **Figure 3.2** yellow patches) and at long asynchrony (Pre SOA 300 > Post SOA 300; **Figure 3.2** red patches, both cluster-size corrected to $\alpha = 0.05$ with starting $p = 0.01$, $t = 2.76$). It should be noted that these regions of overlap lie in canonically-identified unisensory and multisensory convergence areas (see **Table 2.1**), the most prominent of which (127 combined voxels) is centered on the right posterior superior temporal sulcus (pSTS), an average of 28.13 (± 4.16 SEM) voxels away from the centers of gravity of pSTS areas of activation identified in prior

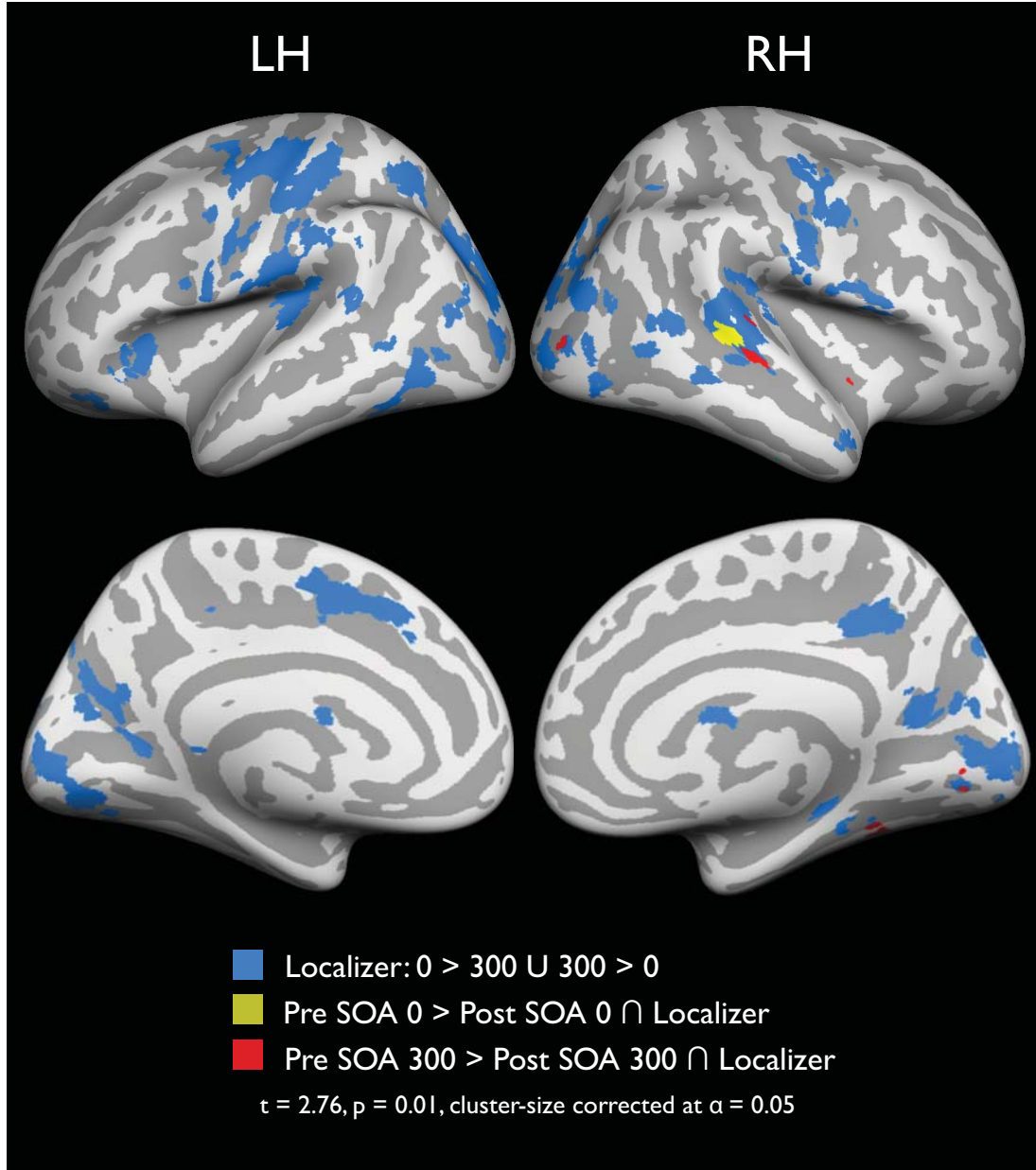


Figure 3.2. Synchrony-responsive networks change with perceptual training. Group data from a random-effects analysis meant to identify regions of cortex that are sensitive to synchrony versus asynchrony of audiovisual stimulus pairs in localizer runs (light blue). Other colors represent clusters that responded differentially to presentation of audiovisual pairs separated by the same stimulus onset asynchrony (SOA) in pre- and post-training scans. Cortical surface is created from the group average. Statistics for all contrasts were cluster-size corrected for multiple comparisons to $\alpha = 0.05$.

Table 1. Regions exhibiting localizer and pre-/post-training differences. Regions listed are identified via a conjunction analysis of synchrony- and training-responsive areas. *P* values listed are reported before and after cluster-wise FDR correction for multiple comparisons.

<i>Region</i>	<i>Cluster Size (voxels)</i>	<i>Center of Gravity</i>			<i>T</i>	<i>p-unc</i>	<i>p-FDR</i>
		<i>X</i>	<i>Y</i>	<i>Z</i>			
<i>Pre 0 > Post 0</i> ∩ <i>Localizer 0>300</i>							
Right Posterior Superior Temporal Sulcus	77	62	-34	-4	3.83	0.003313	0.00994
<i>Pre300Post300</i> ∩ <i>Localizer 0>300</i>							
Right Medial Superior Cerebellum (Culmen)	162	36	-60	-32	3.34	0.007434	0.029555
Right Posterior Superior Temporal Sulcus	50	52	-22	-10	3.23	0.008987	0.029555
Right Cuneus	3	4	-76	2	3	0.013434	0.029555
Left Superior Temporal Gyrus	2	-46	-34	12	2.89	0.016233	0.029761

audiovisual synchrony detection studies (Calvert et al., 2000; Calvert et al., 2001; Dhamala et al., 2007; Noesselt et al., 2007; Stevenson et al., 2010) . Using these overlapping areas as cortical regions of interest (ROIs), we are able to analyze the details of their activity changes over the training period.

BOLD activity in pSTS decreases with training

The pSTS ROIs identified by this conjunction analysis make up a 127-voxel region in middle-posterior STS consisting of separate but adjacent areas that exhibit significant decreases in BOLD activation after training (**Fig. 3.3a** – colors the same as for **Fig. 3.2**). To quantify these differences, the mean event-related ROI time courses were extracted for each individual, and the degree of BOLD activity change was taken as the mean percent signal change. The greatest magnitude change was seen in the simultaneous (SOA 0) and highly asynchronous (SOA 300) conditions (**Fig. 3.3b-d**). In contrast, little change was observed for the intermediate, individually-defined right window size (RWS) condition in either the anterior pSTS ROI (**Fig. 3.3b**; SOA 0: $t_{10} = 3.09$, $p = 0.0115$; SOA 300: $t_{10} = 1.54$, $p = 0.155$), the posterior pSTS ROI (**Fig. 3.3d**; SOA 300: $t_{10} = 2.57$, $p = 0.0278$), or the two taken as one combined ROI (**Fig. 3.3c**; SOA 0: $t_{10} = 2.37$, $p = 0.0393$; SOA 300: $t_{10} = 1.989$; $p = 0.0748$).

BOLD signal decreases like those seen here have been interpreted to reflect improved efficiency of processing, wherein neuronal firing shifts from a large population of neurons to a more specialized subset (Mukai et al., 2007). Thus, the direction of change after training could be interpreted to support the hypothesis that training increases the efficiency of processing of objectively simultaneous and highly asynchronous

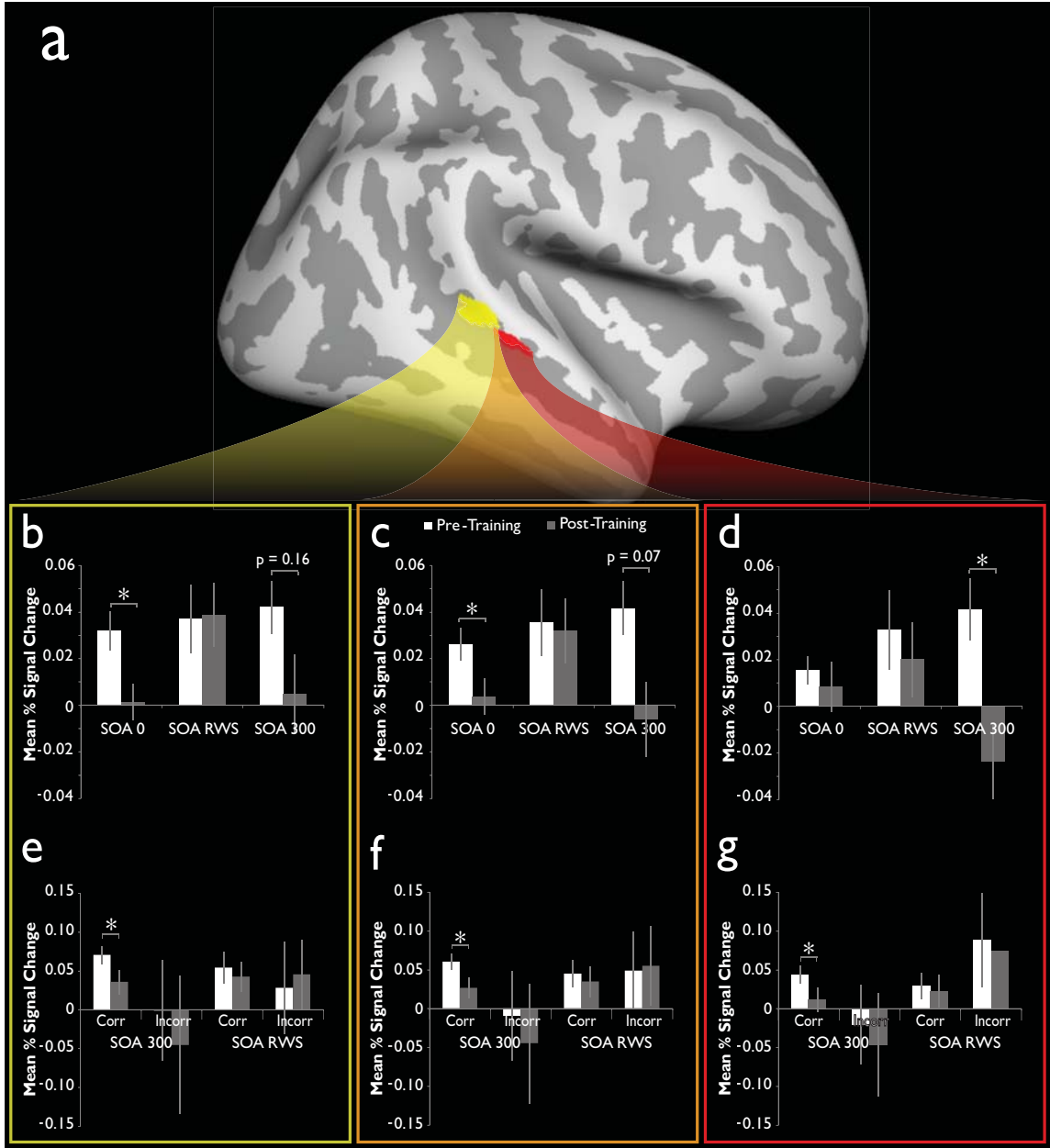


Figure 3.3. BOLD activity in posterior superior temporal sulcus (pSTS) decreases with training.
a. The pSTS regions of interest (ROIs) as identified in Figure 3.2. **b-d.** Mean percent signal change for all voxels in the SOA 0 ROI (**b**, yellow box), the SOA 300 ROI (**d**, red box), and the two combined (**c**, orange box) over the course of the extracted FIR time course. Significant decreases are found at these SOAs but not at the intermediate (SOA RWS) interval. **e-g.** Mean percent signal change as a function of trial accuracy for SOA 300 and SOA RWS trials, for the SOA 0 (**e**), SOA 300 (**g**), and combined (**f**) ROIs *, $p < 0.05$.

stimulus pairs. In order to test this hypothesis further, time courses modeled on correct and incorrect (i.e., committed error) trials were extracted from the pSTS ROIs and analyzed for changes over the course of training. The expectation for these analyses with an increased efficiency of processing framework is a significant decrease in activation for correct SOA 300 trials, but not for correct SOA RWS trials or for incorrect trials at either SOA. This pattern of results was indeed observed in both of the individual ROIs (**Fig. 3.3e** and **3.3g**; SOA 0: $t_{10} = 3.04$, $p = 0.0125$; SOA 300: $t_{10} = 1.90$, $p = 0.0873$), as well as the combined pSTS ROI (**Fig. 3.3f**; $t_{10} = 2.71$, $p = 0.022$). It may also be noted that incorrect trials in all three ROIs elicited greater mean BOLD activity in the RWS trials than in the SOA 300 trials, although none of these values reached statistical significance.

Auditory and visual cortices as well as superior cerebellum exhibit decreases in BOLD activity after training

In addition to the changes in activity observed in pSTS, similar training-related decreases were seen in visual and auditory cortices for the SOA 300 condition. **Figure 3.4a** depicts the region of overlap between the auditory localizer and this training-induced ROI. The region lies on posterior border of BA41, roughly corresponding to the junction between primary and secondary auditory cortex on probabilistic cytoarchitectonic maps (Rademacher et al., 2001). Quantification of the difference in activity observed in this region reveals a similar pattern of change to that seen in the pSTS ROIs at the SOA 300 condition ($t_{10} = 2.72$, $p = 0.0214$), but little to no change on either of the other conditions (**Fig. 3.4b**). Also similar to the results seen in the pSTS ROIs, activity in this region declines selectively for correct trials on the SOA 300

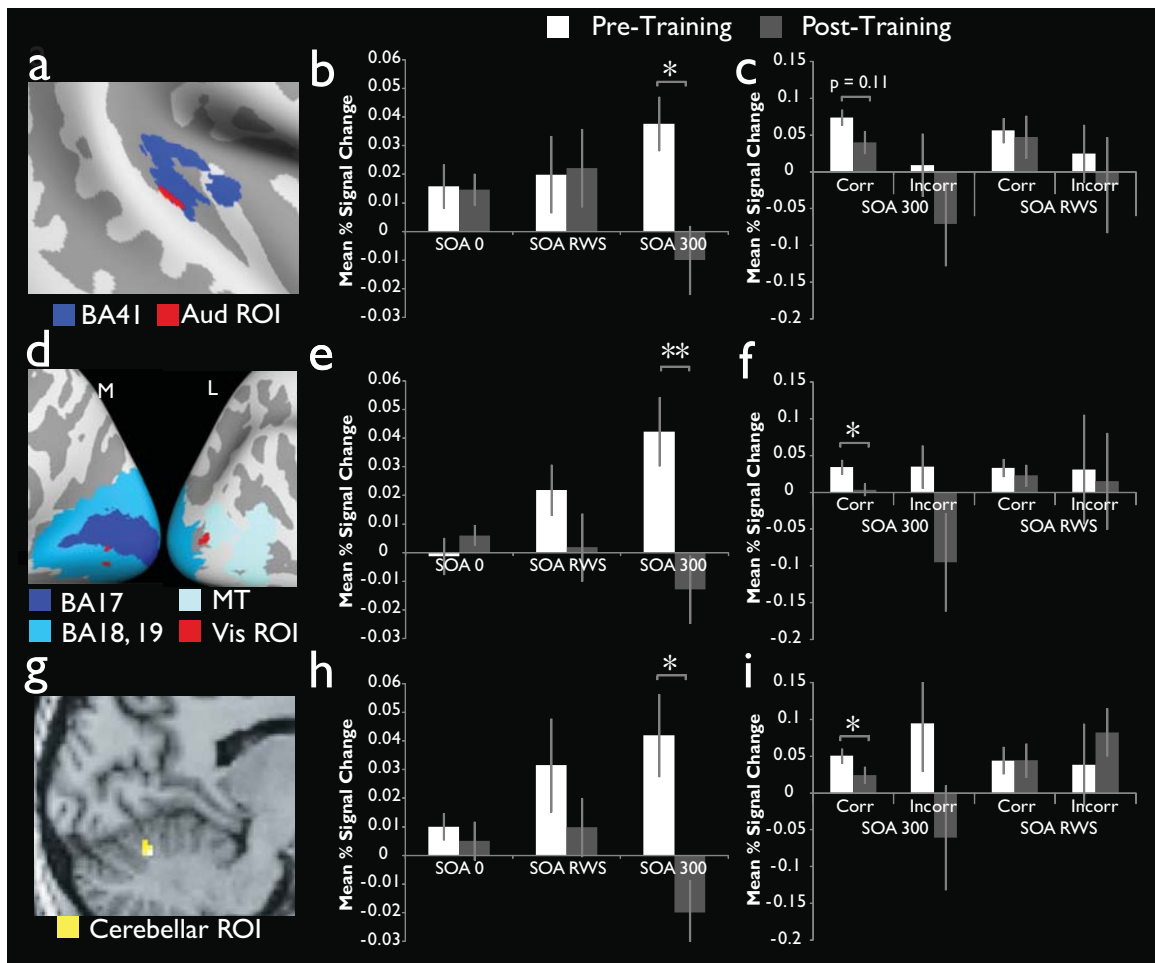


Figure 3.4. Auditory and visual cortices as well as superior cerebellum exhibit decreases in BOLD activity after training. *a.* Auditory ROI (red) and its physical relationship to BA 41 (dark blue). The ROI lies ventral/posterior border of BA 41, corresponding to auditory belt regions. *b.* Mean percent signal change in this region for the three SOA conditions. *c.* Mean percent signal change as a function of trial accuracy for this region. *d.* Visual ROIs (red) and their physical relationship to BA 17, BA 18, BA 19 and MT. The ROIs lie in all three of these areas in addition to an intermediate area, most likely V4. *e.* Mean percent signal change in the visual ROI for the three SOA conditions. *f.* Mean percent signal change in the visual ROI as a function of trial accuracy. *g.* Cerebellar ROI (yellow) in parasagittal section [$x = 36$]. The ROI is located in right vermal lobules VI and VII. *h.* Mean percent signal change in the cerebellar ROI for the three SOA conditions. *i.* Mean percent signal change in the cerebellar ROI as a function of trial accuracy. *, $p < 0.05$; **, $p < 0.01$.

condition only, although this drop does not reach statistical significance (**Fig. 3.4c**; $t_{10} = 1.713$, $p = 0.117$).

A similar but more pronounced pattern of results is seen in visual cortex. **Figure 3.4d** illustrates the overlap-defined regions of interest in visual areas in relation to Brodmann areas 17 through 19 on the medial aspect of the occipital lobe, and for the middle temporal area (MT) on its lateral aspect. The two smaller areas of activation displayed on the medial aspect of the occipital lobe appear to lie within BA 17 and 18 (MNI305 template). Once again, it should be noted that these regions of interest represent the areas of overlap between the temporal localizer (SOA 0 > SOA 300) and regions that change following training (i.e., Pre SOA300 > Post SOA300), which have themselves been statistically corrected via cluster-size thresholding; thus, these clusters, while small, do represent significant activations. The larger area of overlap on the lateral surface (40 total voxels) appears to lie at the border between MT and the lateral borders of BA 18/19 in the middle occipital sulcus. Combined analysis of BOLD activity in this regions reveals a large decrease after training in SOA 300 conditions (**Fig. 3.4e**; $t_{10} = 3.17$, $p = 0.0099$), as well as a significant preferential decrease in activity in SOA 300 correct trials (**Fig. 3.4f**; $t_{10} = 2.80$, $p = 0.0188$).

Of the group-defined regions of interest listed in **Table 2.1**, the activation found in the superior cerebellum stands out as the only non-cortical area. **Figure 3.4g** shows a parasagittal section of the cerebellum, with the cerebellar ROI situated in the vermal declive (AAL Region VI (Schmahmann et al., 1999; Tzourio-Mazoyer et al., 2002)). Again consistent with the previously described patterns, quantification of BOLD activity during pre- and post-training conditions revealed a significant decrease only in the SOA

300 condition (**Fig. 3.4h**; $t_{10} = 2.99$, $p = 0.0135$). Also similar to the results seen in the previously characterized ROIs, activity in this region declines preferentially in the correct trials on the SOA 300 condition only (**Fig. 3.4i**; $t_{10} = 2.46$, $p = 0.0338$).

Resting state functional connectivity increases are seen between pSTS and auditory areas after training

Having identified a network of cortical sensory areas that appear to play a role in the training-induced changes in the multisensory temporal binding window, we next sought to identify changes in inherent functional coupling between the pSTS ROI and the other nodes in this network. Data collected during pre- and post-training resting-state functional runs were analyzed using the combined pSTS ROI as a seed. **Figure 3.5** highlights those cortical areas that increase their resting state functional coupling with pSTS after training (**Fig. 3.5a**). These include right secondary auditory cortex (50 voxels; peak intensity $t = 5.36$), right parahippocampal gyrus (69 voxels; peak intensity $t = 8.77$), inferior parietal lobule (29 voxels; peak intensity $t = 5.81$), and left premotor cortex (13 voxels; peak intensity, $t = 4.26$). In addition, increased resting state functional coupling after training was observed in the superior colliculus (44 voxels; peak intensity $t = 6.53$; **Fig. 3.5b**) as well as superior cerebellum (88 voxels; peak intensity $t = 6.22$). In contrast, a test for areas that decreased their coupling with the pSTS ROI after training yielded no significant results.

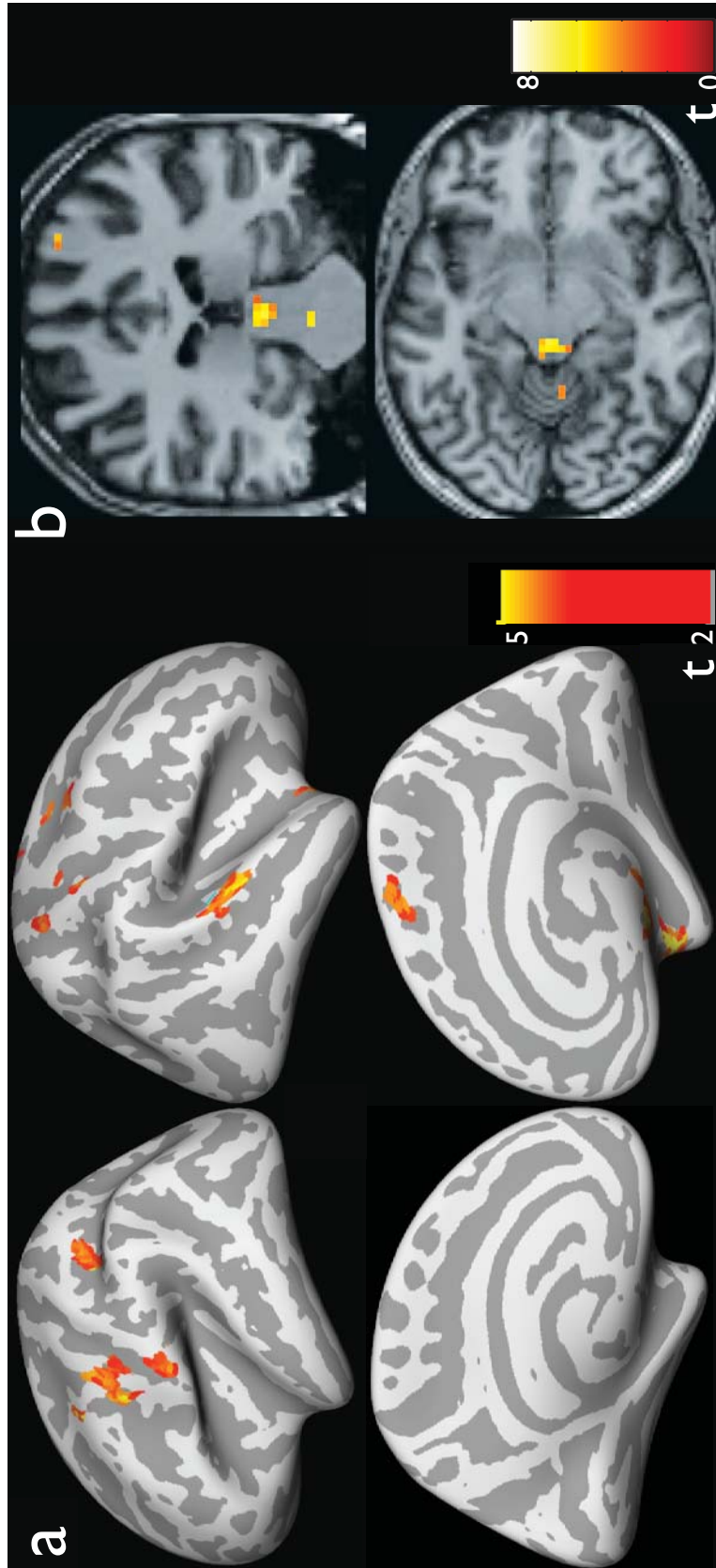


Figure 3.5. Resting state functional connectivity increases are seen between pSTS and auditory areas after training. *a.* Cortical areas whose activity is more positively correlated with the combined pSTS ROI after training. *b.* Resting state functional connectivity increases between pSTS and superior colliculus after training. Coordinates: [0, -31, -11]. Both images, $t = 2.76$, $p = 0.01$; cluster-size corrected to $\alpha = 0.05$.

Effective connectivity increases between pSTS and visual areas are seen after training

While resting-state functional connectivity results indicate alterations in inherent functional connectivity between the pSTS and an array of cortical and subcortical structures (including auditory cortex) following perceptual training, they are not able to reveal task-related or directional connectivity changes between these (and other) areas. Hence, as the next step in these analyses, a dynamic causal modeling approach was taken in order to examine training-related changes in effective connectivity. **Figure 3.6a** illustrates a dynamic causal model based upon the unisensory and multisensory ROIs identified in the current study, the connections (statistical dependencies) among which are based upon previous functional (Noesselt et al., 2007) and neuroanatomical (Barnes and Pandya, 1992; Hackett et al., 1998; Cappe and Barone, 2005) studies. The model hypothesizes both feedforward and feedback effective connectivity between pSTS and visual and auditory cortices. Placed into the context of a blocked design during which auditory-alone (A), visual-alone (V) and combined audiovisual (AV) stimulus pairs were presented, these stimulation periods (interspersed with rest blocks) serve as predictable direct driving forces for the visual and auditory areas being examined (**Fig. 3.6a**, red arrows).

Working from this simple model and guided by the knowledge that simultaneity perception has been shown to be accompanied by increases in effective connectivity among the nodes of this network (Noesselt et al., 2007), we hypothesized that the proposed effective connections between pSTS and visual and auditory cortices might be modified after training (**Fig. 3.6b**, models 1-4; for example, model 1 [leftmost panel]

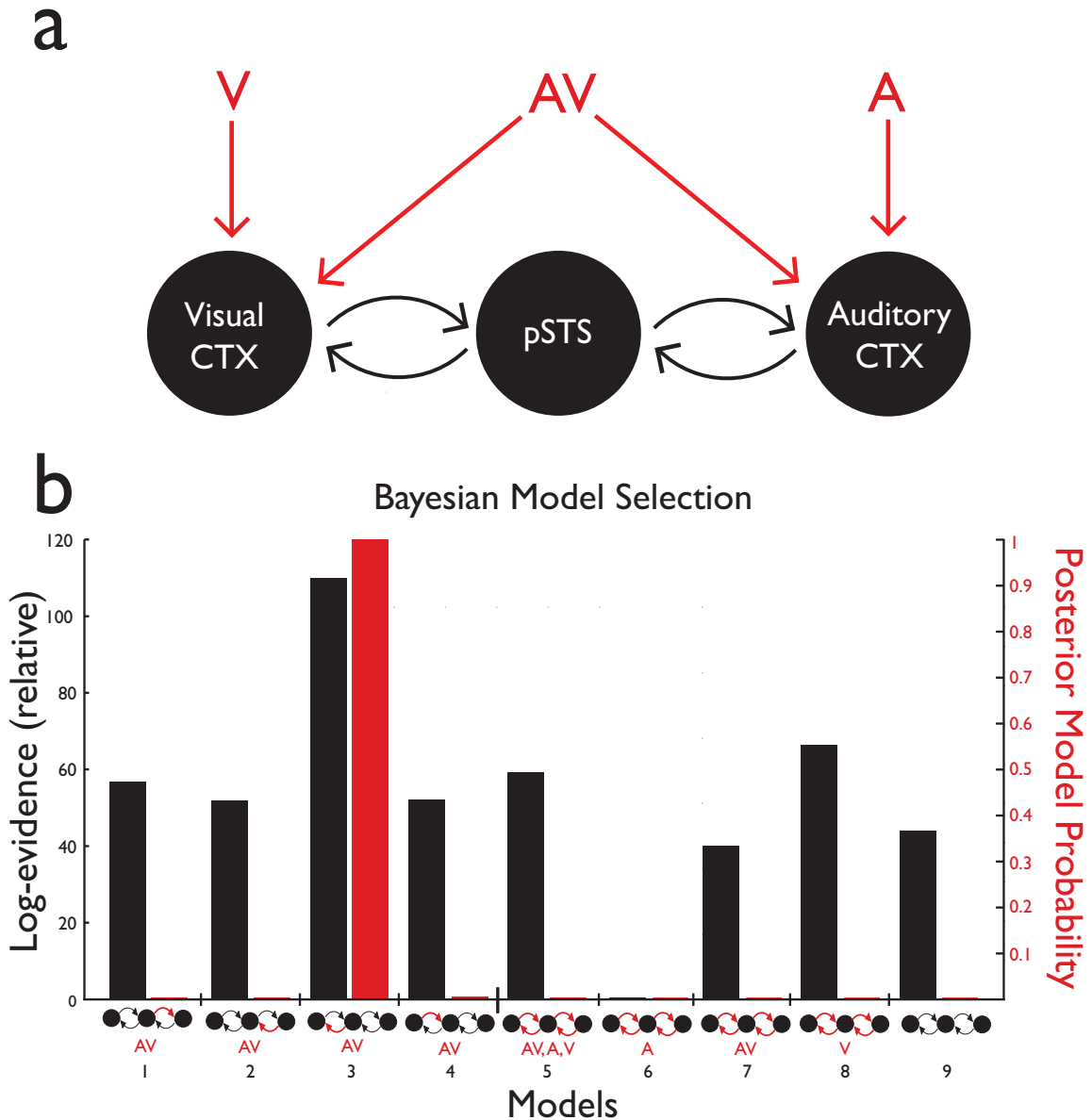


Figure 3.6. Effective connectivity changes between pSTS and visual areas are observed after training. *a.* Model demonstrating proposed interconnectivity between the pSTS, visual, and auditory ROIs with stimulation during the blocked-design protocol. Red arrows indicate driving forces. *b.* Bayesian model selection for each of the models illustrated. In the models, red arrows indicate connections that are modulated with training and A, V, and AV indications indicate the conditions wherein connectivity is proposed to be modulated. A, auditory only; V, visual-only; VA, visual-auditory presentation.

captures changes between pSTS and auditory cortex, but only in the feedback domain). The models corresponding to each of these alternatives, along with additional controls modeling null alternatives (overall change in all conditions, models 5-8, as well as no modulation from pre- to post-training, model 9) were analyzed via Bayesian model selection. In this probabilistic framework, the evidence that a given model has produced the data observed is calculated as the product of the likelihood of observing the data given such a model and the prior likelihood that the model is true (Chater and Oaksford, 2008; Rosa et al., 2010). This value was estimated for each model in each individual data set, and a group Bayes factor (i.e., the posterior probability that the data were generated by one model relative to another) was calculated for all subjects, relative to the model garnering the least evidence. The bar graphs shown in **Figure 3.6b** depict the results of this analysis. The model advocating changes in effective connectivity between pSTS and visual cortex after training (i.e., model 3) garners the most log-evidence (110 relative units), followed most closely by the model predicting that all functional connections among the different ROIs would be modulated after training during visual presentation (i.e., model 8; 66 relative units). As a difference of only 3 units of log-evidence corresponds roughly to a posterior model probability of 95% for the superior model (Stephan et al., 2007), a difference of 44 units here corresponds to a posterior probability of virtually 1.00 for model 3 (**Fig. 3.6b**, red bars), indicating that changes in effective connectivity from pSTS to visual cortex after training are most likely to have produced the data observed.

Discussion

The data reported here represent the first attempt to both characterize and alter the dynamics of a network of sensory areas underlying a fundamental and ethologically important perceptual process – the perception of audiovisual simultaneity. In doing so, we have illustrated not only that certain nodes of that network are strongly tied to the perception of simultaneity, but also that their activity and interrelatedness change after perceptual learning in a way that clarifies the roles they play during everyday perception of multisensory events.

That the neural signature of learning in this study was associated with a decrease in BOLD activity at key sensory nodes is not unprecedented; activity decreases in sensory cortices have been shown to correlate with perceptual learning of visual contrast (Mukai et al., 2007), illusory contours (Maertens and Pollmann, 2005), and orientation discrimination (Schiltz et al., 1999), along with implicit learning of visual categories (Reber et al., 1998), as well as following training on a visual object tracking task (Tomasi et al., 2004). A consistent interpretation of these BOLD decreases has been that they reflect improved efficiency of processing, manifested as a shift in neuronal firing from a large population of neurons to a smaller, more specialized (and perhaps more sparsely distributed) subset (Mukai et al., 2007). Such an effect is consistent with electrophysiological studies of visual recognition memory (Miller et al., 1991) and perceptual learning (Ghose et al., 2002), as well as with most efforts to model the plasticity of local neural architecture giving rise to perceptual learning (Hoshino, 2004).

Given that adaptation to repeated sensory stimuli also produces decreases in cortical activity profiles (Desimone, 1996), it is important to distinguish between adaptation and perceptual learning as the likely agent of the changes observed here. The fact that the BOLD decreases were specific to the objectively simultaneous (i.e., SOA 0) and highly asynchronous (i.e., SOA 300) conditions, and not to the individualized threshold asynchrony condition (i.e., SOA RWS), argues strongly in favor of perceptual learning, since adaptation effects should accompany all conditions equivalently and supports the hypothesis that increased processing efficiency (and thus, decreases in effort needed to reach maximal performance on less challenging conditions (Schiltz et al., 1999)) is driving the observed changes. Moreover, the fact that the decreases were preferentially associated with correct trials also argues strongly for a learning-based interpretation of the data.

As striking as the training-related decreases in BOLD activity observed at these loci were the changes in functional connectivity seen between brain areas. Analysis of resting state functional connectivity data was used to examine changes in inherent functional coupling between the pSTS ROI and other regions. By contrast, dynamic causal modeling using data acquired during stimulus presentation revealed the ways in which interactions between important nodes changed after training. Key results of these connectivity-based analyses include that pSTS exhibits a higher degree of inherent (i.e., resting state) functional coupling with auditory cortex after training, and that pSTS has an increased influence on activity in visual cortex after training, but only during combined audiovisual stimulus presentation. This combination of evidence suggests a model in which pSTS becomes the director of multimodal temporal processing improvements

through enhanced coupling with unisensory visual and auditory areas. Thus, greater functional coupling between pSTS and visual cortex after training leads to improvements in visual processing and heightened audiovisual temporal discrimination ability. Anatomical and electrophysiological studies of pSTS suggest that such a role is plausible: primate auditory belt regions exhibit strong feedforward and feedback interconnectivity with pSTS (Barnes and Pandya, 1992; Hackett et al., 1998; Cappe and Barone, 2005); likewise, pSTS receives abundant input from adjacent visual areas (Barnes and Pandya, 1992), and projections from pSTS to lateral occipital visual areas and even primary visual cortex have been well characterized (Falchier et al., 2002). Visual and auditory response latencies in pSTS (Bruce et al., 1981) would allow ample time for feedback to influence visual processing, whether by direct mechanisms or by phase resetting of neuronal oscillations (Lakatos et al., 2007) between the two regions. In fact, pSTS has already been implicated as a driving source in numerous studies of cross-modal influences on visual (Watkins et al., 2006; Noesselt et al., 2007; Watkins et al., 2007) and auditory (Kayser et al., 2007) cortices. The current data greatly expand this framework by providing the first evidence that pSTS plays a central role in orchestrating the plastic changes that accompany multisensory perceptual learning.

The enhanced resting state functional connectivity between pSTS and both superior colliculus (SC) and superior cerebellum is interesting because of the roles these regions play in multisensory integration and supramodal temporal processing, respectively, but whether and how their alteration factors into the behavioral changes observed remain unresolved. The SC, despite its central role in multisensory integration, and its activation in several studies examining audiovisual simultaneity perception, does

not itself exhibit BOLD changes after training. In contrast, the SC does show enhanced coupling with pSTS after training, and offers support to the view that connectivity of SC with multisensory and unisensory cortical areas are important for audiovisual simultaneity detection (Bushara et al., 2001; Dhamala et al., 2007). The presence of activation changes and functional interactions between the pSTS and the cerebellum carries with it implications for alterations of supramodal timing mechanisms via audiovisual perceptual training, although vermal and hemispheric lobule VI has very recently been shown to be directly responsive to visual and auditory stimuli (Baumann and Mattingley, 2010), as well as in the anticipation of sensory events (Buetti et al.).

From a clinical perspective, the similarity between the regions influenced by the current perceptual training protocol and those altered in individuals with disorders of multisensory processing is striking. For example, in addition to having enlarged multisensory temporal binding windows (Hairston et al., 2005), individuals with developmental dyslexia also exhibit altered BOLD activity in pSTS compared to typical-reading controls during auditory and visual non-word rhyming (Rumsey et al., 1992; Shaywitz et al., 2002), pseudoword processing (Brunswick et al., 1999; Paulesu et al., 2001), and letter-word sound matching (Blau et al., 2010). Additionally, an entire school of thought based on findings of functional deficits in children with dyslexia (Fawcett et al., 2001; Stoodley et al., 2005; Stoodley and Stein, 2009) has emphasized the role of cerebellum in development of reading automaticity. Supporting evidence for this comes from demonstrated BOLD decreases in vermal lobule VI in children with dyslexia as compared with typical-reading controls during pattern learning (Jenkins et al., 1994). Similarly, in children with autism spectrum disorders (ASD), pSTS appears to be an

important affected node, with decreases in gray matter concentration, resting hypoperfusion and abnormal activation all being noted (Zilbovicius et al., 2006; Redcay, 2008). Moreover, amodal (Brock et al., 2002) and multisensory (Foss-Feig et al., 2010) timing deficits have been demonstrated in individuals with ASD, and may be associated with morphological (Courchesne et al., 1988; Courchesne, 1995; Mitchell et al., 2009) and functional (Mostofsky et al., 2009) changes in the superior cerebellum that correlate with disease severity. Lastly, it is clear that multisensory processing (Ross et al., 2007; de Jong et al., 2009; Jardri et al., 2009) and pSTS function in the context of audiovisual speech binding (Szycik et al., 2009) is altered in individuals with schizophrenia when compared with matched typically-developing controls. Given that these deficits and the plastic changes demonstrated here are likely to be reflective of changes in local neural architecture (Brock et al., 2002), these results point to multisensory temporal training as an especially promising new avenue of exploration for remediation of these disorders.

Overall, the results reported here illustrate that the narrowing of the multisensory temporal binding window is accompanied by decreases in BOLD signaling between a network of multisensory and unisensory areas centered on the pSTS, as well as by changes in resting-state and task-related functional coupling among these areas. These data indicate that the influence of multisensory cortical regions on unisensory processing may be a driving factor behind these improvements in multisensory temporal processing.

Materials and Methods

Subjects

Thirteen right-handed Vanderbilt students and employees (mean age = 23.4 years; 8 female) participated in both the behavioral and imaging portions of the study. All participants had self-reported normal hearing and normal or corrected-to-normal vision. None had any personal or close family history of neurological or psychiatric disorders, and all completed fMRI screening and informed written consent form. Data from two subjects were discarded before the analysis phase due to either inability to perform the behavioral task or experimenter error. All procedures were approved by the Vanderbilt University Institutional Review Board (IRB).

Procedure

Experimental procedures took place over the course of two days. Participants first took part in a behavioral assessment, followed by a pre-training MRI scan. They then underwent 1 hour of behavioral training outside the scanner and underwent a post-training behavioral assessment. Each participant then returned to the laboratory the next day for a final behavioral assessment and a final MRI scan. The details of each of these procedures are outlined below.

Behavioral

Behavioral Assessment

Before training began, participants engaged in a two-interval forced choice (2-IFC) audiovisual simultaneity judgment task (**Fig. 3.1**). In the task, participants

determined which of the two audiovisual stimulus pairs was presented synchronously (stimulus onset asynchrony, SOA, of 0 ms) by responding 1 or 2, respectively, on a standard computer keyboard. Participants sat in a dark and sound-attenuated room, 48 cm from a computer monitor (NEC MultiSync FE992; resolution: 800 x 600 pixels; vertical refresh rate: 120 Hz). MatLab version 7.7.0.471 2008b (The Mathworks, Inc., Natick, MA) equipped with PsychToolbox 3.0 was used for stimulus and protocol control as well as to acquire all behavioral data. Temporally precise presentation of auditory and visual stimuli was achieved with the use of high-performance video (ATI Radeon HD 2400 Pro) and audio (Creative SoundBlaster X-FI) cards, the latter of which was equipped with ASIO drivers for use in the PsychToolbox to ensure low-latency and high-fidelity presentation of auditory stimuli.

A white crosshair fixation marker (1 cm x 1 cm) appeared on a black background for the duration of each trial. The visual stimulus consisted of a white annulus on a black background subtending 15° of visual space, with an outer diameter of 12.4 cm and an inner diameter of 6.0 cm (area = 369.8 cm^2). This stimulus was presented for one refresh duration on the NEC monitor above (refresh rate: 60 Hz; one refresh duration: 17 ms; stimulus was present on screen for 13 ms).

The auditory stimulus was an 1800 Hz tone burst and was 17 ms in duration. The stimulus was presented to both ears via supraaural headphones (Philips SBC HN110) with no inter-aural time or level differences. The tone burst was calibrated with a Larson-Davis sound level meter (Model 814). Acoustic stimuli were presented at 110.4 dB SPL un-weighted using impulse detection and flat weighting settings.

The visual and auditory stimuli were presented at SOAs ranging from -300 ms (auditory stimulus leading visual) to +300 ms (visual stimulus leading auditory) at 50 ms intervals. SOAs were verified externally with an oscilloscope, and stimulus presentation times recorded internally in MatLab were adjusted to reflect the presentation times observed on the oscilloscope. This allowed the script to record an estimate of what the SOA for each trial was as if verified externally. By this estimate, presentations were on average presented within an error of 2.54 ms of the desired SOA (standard deviation = 0.37 ms among subjects). This error was the best possible given the inherent imprecision in the response times of the stimulus presentation hardware. The task consisted of 325 total trials (25 cycles x 13 trials/cycle). By nature of the 2-IFC task structure, simultaneous and non-simultaneous pairs were presented with equal likelihoods.

Behavioral Training

The behavioral training portion of the study used the same stimuli as the assessment, with the exception that the subject was presented with either the phrase “Correct!” paired with a yellow happy face, or “Incorrect” paired with a blue sad face after each trial, corresponding to whether they answered correctly or not. These faces (area = 37.4 cm²) were presented at the center of the screen for 0.5 s. Training consisted of three blocks of 120 trials (20 cycles x 6 trials/cycle) and consisted of SOAs from -150 ms to 150 ms. The total training session lasted for about 45 minutes.

Imaging

Procedure

The scanning protocol consisted of anatomical scans followed by a blocked-design multisensory stimulation protocol, two event-related runs, a resting state functional connectivity scan, and three final event-related runs. The scanning procedure was the same for both days. A 3-Tesla fMRI scanner (Philips Intera Achieva 3T) was used to generate high-resolution anatomical images and to measure blood-oxygen-level-dependent (BOLD) signals. The high-resolution anatomical scans produced 3D T1-weighted images (170 slices, TR = 7.982 ms, TE = 3.68 ms, thickness = 1 mm, field of view (FOV) = 256 x 256 mm², matrix size = 256 x 256 x 170; voxel size = 1x1x1 mm). The event-related runs were acquired with a T2*-weighted standard EPI sequence (33 slices acquired in an ascending interleaved pattern, TR = 2.5 s, TE = 35 ms, slice thickness = 4 mm, FOV = 240 x 131.5 mm², matrix size = 80 x 78 x 33; voxel size = 3x3x4 mm, 0.5 mm gap inclusive) for a duration of 5 minutes and 57 seconds per run; full-brain coverage was achieved with the use of a full-volume (birdcage) RF head coil, and 135 volumes were acquired per run.

Visual stimuli were presented via a projector (NEC Model MT1050; resolution: 800 x 600 pixels; refresh rate = 60 Hz) onto a frosted glass screen in the scanning room that was viewed by participants via a front-projection MR-compatible mirror mounted on the head coil. Acoustic stimuli were presented with intra-aural, MR-compatible insert earphones (Sensimetrics MRI-Compatible Insert Earphones Model S14) and tips (Hearing Components Inc. Comply Canal-Tips). Super-aural sound-attenuating headphones were placed over them to curb effects of scanner noise. Audibility and

visibility of stimuli were checked against pre-set marks before and after each scan for each subject to ensure standardization of stimulus presentation during each imaging session.

Directly after the completion of anatomical image acquisition, a functional run was completed with the use of the same visual and auditory stimuli in a blocked design. Blocks consisted of twenty presentations of either auditory-only (A), visual-only (V), or combined visual-auditory (VA) stimuli over the course of ten volume acquisitions. In a manner similar to that used during the event-related runs, a standard EPI sequence was used to acquire BOLD data (170 volumes; TR: 2000 ms; TE: 35 ms; FOV: 240.000 mm x 131.500 mm) during the course of this 5-min, 56 s run. A, V, and VA blocks were pseudo-randomized and counterbalanced across subjects, and a rest block was interposed after every third block. During this run, participants were instructed to remain still and pay close attention to the stimuli.

The physical characteristics of the stimuli used during the event-related runs were identical to those used in the behavioral portion of the study except that only three SOA conditions were used: 0 ms (simultaneous), 300 ms, and a third SOA determined by participants' baseline behavioral assessment, defined as the size of that individual's right window size (RWS; see Data Analysis for details of window size measurement). If the RWS exceeded 300 ms, then the SOA was set to 150 ms (true for 2 subjects).

Participants performed the 2-IFC task as they had during the behavioral assessment. They were asked to respond as quickly and accurately as possible. Within the 2.5-s TR, presentation of visual and auditory stimuli (17 ms each) and an 800-ms inter-stimulus interval (ISI) allowed for 1.3-1.6 s to respond, depending on SOA. During each event-

related run, presentation of these events was temporally jittered and counterbalanced with the use of an 127-event m-sequence(Buracas and Boynton, 2002) that ended with a 25-s period of no stimulus presentation save the fixation cross.

During the resting-state functional connectivity scans, participants were instructed to close their eyes and think neutral thoughts. The data acquisition protocol used in the blocked-design multisensory stimulation run was also used for data collection here.

Data Analysis

Estimation of Window Size

All behavioral data were stored in individual-subject .mat files for use in analysis with MatLab. Individual subject raw data were used to calculate mean accuracy at each SOA for all assessments. Mean data from each individual were fit with two sigmoid curves generated using the MatLab *glmfit* function, splitting the data into left (auditory presented first) and right (visual presented first) sides and fitting them separately. The criterion at which to measure the breadth of the temporal window was equal to half the distance between individuals' lowest accuracy point at baseline assessment and 1 (~ 75% accuracy). These criteria were then used to assess the breadth of the distributions produced by each individual's assessment data throughout the duration of the training period(Powers et al., 2009). Distribution breadth was then assessed for both the left side (from zero to the left-most point at which the sigmoid curve crossed the criterion line) and the right side (from zero to right intersection point) and then combined to get an estimation of total distribution width. This measure was then used as a proxy for the size of each individual's window at each assessment, and the right-sided value was used as the right window size (RWS) SOA to be used in both scanning sessions. An example of

the result of this process may be seen in **Figure 3.1c**. It should be noted that when mean data from any individual assessment were unable to be fit with a sigmoid curve all data from this individual were discarded. Analysis of differences in window size across time was conducted by performing three paired-samples t-tests (Bonferroni corrected for multiple comparisons) to assess for differences in mean window size across the four assessments.

Imaging Data Preprocessing

Imaging data from each run were stored as Philips .par and .rec files, which were converted to NIFTI (.nii) files at the start of preprocessing. Statistical Parametric Mapping version 8 (SPM8) was used for the entirety of imaging data preprocessing and statistical analysis. Functional data were corrected for slice acquisition timing (reference slice = 1) and motion during the scanning session (registered to mean), and resliced before being coregistered with the high-quality anatomical data from that session. Both anatomical and functional images were then warped into standard Montréal Neurological Institute (MNI) space with 2 mm isometric voxels using trilinear interpolation settings. Finally, functional images were spatially smoothed (FWHM of Gaussian smoothing kernel: 8x8x8 mm) and readied for statistical analysis.

Statistical Analysis

By virtue of the event-related design utilized in this study, a Finite Impulse Response (FIR) basis set was used to model the BOLD response (microtime resolution: 16 ms; microtime onset: 1; high-pass filter: 128 Hz; window length: 25 s; 10 basis

functions; event durations: 0 ms) with conditions corresponding to presentation of each SOA (0 ms, RWS, 300 ms), as well as to response-related parameters (correct/incorrect responses, errors of commission and omission) in both separate and combined models.

To identify regions of cortex that are differentially active to synchrony versus asynchrony of presented audiovisual pairs, we isolated four event-related runs (runs 4 and 5 of both pre- and post-training scanning sessions) for use as a localizer of these cortical regions, combined into a single model. Contrasts and statistical parametric maps were generated using canonical weighting of the FIR time bins. Resulting maps were generated on an individual basis and also served as the basis for random effects group analysis with predictors separated for each subject. For this analysis, cluster-size thresholding (Forman et al., 1995) was used for correction (SPM extension *CorrClusTh*, using the Newton-Raphson search method). With a starting voxel-level threshold of 0.01 ($t = 2.76$) uncorrected, this resulted in a cluster level of 161 mm³, corresponding to a corrected false-positive probability of 5% or less. The remaining 6 runs (3 pre-training, 3 post-training) were included in a second model for analysis of pre-/post-training changes. Population-level inferences were based on a second-level paired-samples random effects analysis. As with the localizer, cluster-size thresholding was employed for correction, resulting in cluster and 544 mm³ and 784 mm³ for the Pre-/Post-Training SOA 0 and SOA300 comparisons, respectively; these corresponded to a corrected false-positive probability of 5% or less.

Active voxels in the group contrast of 0 ms > 300 ms for the localizer runs were identified as synchrony-sensitive areas, and within-condition (SOA), between treatment (pre- vs. post-training) contrasts were then generated. Synchrony-sensitive regions

exhibiting altered activity profiles between the pre- and post-training sessions were identified in a conjunction analysis and defined as regions of interest (ROIs). For clarity of visualization, statistical maps were taken from SPM8 and projected onto a custom surface made up of averaged cortical surfaces from each individual created in FreeSurfer v. 4.5.0 (Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA) via the SPM *surfrend* extension. When used, identification of Brodmann areas on this custom surface was accomplished via projection of the *fsaverage* MNI305 template (first MNI template, 305 subjects) onto the custom surface via FreeSurfer command *mri_label2label*. Additional identification and breakdown of clusters by anatomical area was accomplished with the use of the WFU_PickAtlas (Lancaster et al., 2000; Tzourio-Mazoyer et al., 2002; Maldjian et al., 2003) within the *xjview* SPM8 extension.

Statistical analysis of these ROIs was undertaken with the use of the MarsBaR (Marseille Boîte À Région d'Intérêt) and REX (Response Exploration for Neuroimaging Datasets) toolboxes, the former of which was used for extraction of FIR average time courses for each cluster (reported in percent signal change), and the latter of which was used to define clusters and generate cluster-wise statistics, correcting to a false detection rate (FDR) of 0.05 or less. Cluster-wise FIR BOLD time course statistical analysis was performed on mean percent signal change values from individual FIR time courses. These values were then compared across the pre- and post-training scans via paired-samples t-test for each stimulus condition.

Resting state data underwent all pre-processing described above, with the addition of a band-pass filter (0.01 – 0.1 Hz) for isolation of slow, low-amplitude fluctuations in

signal. Analysis of functional coupling (Biswal et al., 1995) (bivariate correlation) was performed with SPM's Functional Connectivity Toolbox (*conn*), using the group-defined ROIs as seeds to generate whole-brain beta maps for each individual's pre- and post-training resting state runs. These were then entered into a second-level analysis (paired-samples t-test) in SPM to determine which voxels exhibited an increase in functional coupling with the seed regions. As with the event-related data, p values were corrected via cluster-size thresholding. With a starting voxel-level threshold of 0.005 ($t = 3.25$) uncorrected, this resulted in a cluster level of 95 mm³, corresponding to a corrected false-positive probability of 5% or less.

Dynamic causal modeling (Lee et al., 2006; Kasess et al., 2009; Stephan et al., 2009a; Stephan et al., 2009b) of neural responses to A, V, and AV presentation blocks during the blocked-design functional runs was implemented through the use of the Dynamic Causal Modeling functionality that is part of the SPM8 suite. The volumes of interest (VOIs) used for this purpose corresponded to the previously-defined auditory (**Fig. 3.4a**), visual (**Fig. 3.4b**), and combined pSTS (**Fig. 3.3**, orange) regions of interest. Driving forces were defined as blocks of visual, auditory, and multisensory stimulation, and bi-directional inherent effective connectivity was hypothesized to exist among the three VOIs. Modulatory influence of training on any/all of these relationships was hypothesized and formed the basis of the different models to be evaluated. Log-evidence (relative to the least-likely model) for each of the models was calculated for individual data sets and formed the basis of group-wise FFX Bayesian model selection.

References

- Bao S, Chen L, Kim JJ, Thompson RF (2002) Cerebellar cortical inhibition and classical eyeblink conditioning. *Proc Natl Acad Sci U S A* 99:1592-1597.
- Barnes CL, Pandya DN (1992) Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 318:222-244.
- Baumann O, Mattingley JB (2010) Scaling of neural responses to visual and auditory motion in the human cerebellum. *J Neurosci* 30:4489-4495.
- Bischoff M, Walter B, Blecker CR, Morgen K, Vaitl D, Sammer G (2007) Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia* 45:578-586.
- Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 34:537-541.
- Blau V, Reithler J, van Atteveldt N, Seitz J, Gerretsen P, Goebel R, Blomert L (2010) Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133:868-879.
- Brock J, Brown CC, Boucher J, Rippon G (2002) The temporal binding deficit hypothesis of autism. *Dev Psychopathol* 14:209-224.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369-384.
- Brunswick N, McCrory E, Price CJ, Frith CD, Frith U (1999) Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: A search for Wernicke's Wortschatz? *Brain* 122 (Pt 10):1901-1917.
- Bueti D, Bahrami B, Walsh V, Rees G Encoding of temporal probabilities in the human brain. *J Neurosci* 30:4343-4352.
- Buracas GT, Boynton GM (2002) Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage* 16:801-813.

Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300-304.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649-657.

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14:427-438.

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur J Neurosci* 22:2886-2902.

Chater N, Oaksford M (2008) *The probabilistic mind: Prospects for Bayesian cognitive science*: Oxford University Press.

Colonus H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci* 16:1000-1009.

Courchesne E (1995) New evidence of cerebellar and brainstem hypoplasia in autistic infants, children and adolescents: the MR imaging study by Hashimoto and colleagues. *J Autism Dev Disord* 25:19-22.

Courchesne E, Yeung-Courchesne R, Press GA, Hesselink JR, Jernigan TL (1988) Hypoplasia of cerebellar vermal lobules VI and VII in autism. *N Engl J Med* 318:1349-1354.

de Jong JJ, Hodiament PP, Van den Stock J, de Gelder B (2009) Audiovisual emotion recognition in schizophrenia: reduced integration of facial and vocal affect. *Schizophr Res* 107:286-293.

Desimone R (1996) Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A* 93:13494-13499.

Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JA (2007) Multisensory integration for timing engages different brain networks. *Neuroimage* 34:764-773.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22:5749-5759.

Fawcett AJ, Nicolson RI, Maclagan F (2001) Cerebellar tests differentiate between groups of poor readers with and without IQ discrepancy. *J Learn Disabil* 34:119-135.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636-647.

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2010) An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res* 203:381-389.

Ghose GM, Yang T, Maunsell JH (2002) Physiological correlates of perceptual learning in monkey V1 and V2. *J Neurophysiol* 87:1867-1888.

Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475-495.

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT (2005) Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res* 166:474-480.

Hoshino O (2004) Neuronal bases of perceptual learning revealed by a synaptic balance scheme. *Neural Comput* 16:563-594.

Innes-Brown H, Crewther D (2009) The impact of spatial incongruence on an auditory-visual illusion. *PLoS One* 4:e6450.

Jardri R, Pins D, Bubrovsky M, Lucas B, Lethuc V, Delmaire C, Vantighem V, Desprez P, Thomas P (2009) Neural functional organization of hallucinations in schizophrenia: multisensory dissolution of pathological emergence in consciousness. *Conscious Cogn* 18:449-457.

Jenkins IH, Brooks DJ, Nixon PD, Frackowiak RS, Passingham RE (1994) Motor sequence learning: a study with positron emission tomography. *J Neurosci* 14:3775-3790.

Kasess CH, Stephan KE, Weissenbacher A, Pezawas L, Moser E, Windischberger C (2009) Multi-subject analyses with dynamic causal modeling. *Neuroimage* 49:3065-3074.

Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27:1824-1835.

Kotani S, Kawahara S, Kirino Y (2003) Purkinje cell activity during learning a new timing in classical eyeblink conditioning. *Brain Res* 994:193-202.

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-292.

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000) Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120-131.

Lee L, Friston K, Horwitz B (2006) Large-scale neural models and dynamic causal modelling. *Neuroimage* 30:1243-1254.

Maertens M, Pollmann S (2005) Interhemispheric resource sharing: decreasing benefits with increasing processing efficiency. *Brain Cogn* 58:183-192.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19:1233-1239.

Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365:350-354.

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215-3229.

Miller EK, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254:1377-1379.

Mitchell SR, Reiss AL, Tatusko DH, Ikuta I, Kazmerski DB, Botti JA, Burnette CP, Kates WR (2009) Neuroanatomic alterations and social and communication deficits in monozygotic twins discordant for autism disorder. *Am J Psychiatry* 166:917-925.

Mostofsky SH, Powell SK, Simmonds DJ, Goldberg MC, Caffo B, Pekar JJ (2009) Decreased connectivity and cerebellar activity in autism during motor task performance. *Brain* 132:2413-2425.

Mukai I, Kim D, Fukunaga M, Japee S, Marrett S, Ungerleider LG (2007) Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J Neurosci* 27:11401-11411.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431-11441.

Paulesu E, Demonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD, Frith U (2001) Dyslexia: cultural diversity and biological unity. *Science* 291:2165-2167.

Perrett SP, Ruiz BP, Mauk MD (1993) Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *J Neurosci* 13:1708-1718.

Powers AR, 3rd, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. *J Neurosci* 29:12265-12274.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 13:669-683.

Reber PJ, Stark CE, Squire LR (1998) Cortical areas supporting category learning identified using functional MRI. *Proc Natl Acad Sci U S A* 95:747-750.

Redcay E (2008) The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci Biobehav Rev* 32:123-142.

Rosa MJ, Bestmann S, Harrison L, Penny W (2010) Bayesian model selection maps for group studies. *Neuroimage* 49:217-224.

Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ (2007) Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr Res* 97:173-183.

Rumsey JM, Andreason P, Zametkin AJ, Aquino T, King AC, Hamburger SD, Pikus A, Rapoport JL, Cohen RM (1992) Failure to activate the left temporoparietal cortex in dyslexia. An oxygen 15 positron emission tomographic study. *Arch Neurol* 49:527-534.

Schiltz C, Bodart JM, Dubois S, Dejardin S, Michel C, Roucoux A, Crommelinck M, Orban GA (1999) Neuronal mechanisms of perceptual learning: changes in human brain activity with training in orientation discrimination. *Neuroimage* 9:46-62.

Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M (1999) Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage* 10:233-260.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. *Nature* 385:308.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14:147-152.

Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2002) Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol Psychiatry* 52:101-110.

Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38:387-401.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009a) Bayesian model selection for group studies. In: Elsevier.

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ (2009b) Ten simple rules for dynamic causal modeling. *Neuroimage* 49:3099-3109.

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of asynchronous audiovisual speech perception. *Neuroimage* 49:3308-3318.

Stoodley CJ, Stein JF (2009) The cerebellum and dyslexia. *Cortex*.

Stoodley CJ, Fawcett AJ, Nicolson RI, Stein JF (2005) Impaired balancing ability in dyslexic children. *Exp Brain Res* 167:370-380.

Szycik GR, Munte TF, Dillo W, Mohammadi B, Samii A, Emrich HM, Dietrich DE (2009) Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. *Schizophr Res* 110:111-118.

Teder-Salejarvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. *J Cogn Neurosci* 17:1396-1409.

Tomasi D, Ernst T, Caparelli EC, Chang L (2004) Practice-induced changes of brain function during visual attention: a parametric fMRI study at 4 Tesla. *Neuroimage* 23:1414-1421.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15:273-289.

Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91:484-488.

Watkins S, Shams L, Josephs O, Rees G (2007) Activity in human V1 follows multisensory perception. *Neuroimage* 37:572-578.

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31:1247-1256.

Zilbovicius M, Meresse I, Chabane N, Brunelle F, Samson Y, Boddaert N (2006) Autism, the superior temporal sulcus and social perception. *Trends Neurosci* 29:359-366.

CHAPTER IV

GENERAL DISCUSSION

Summary of Results

The data reported in this volume represent the first evidence that the temporal window of multisensory binding is capable of being narrowed with perceptual training. This change was rapid, occurring over the course of 1-2 hours of training, robust, amounting to a 40% reduction from its original size, and stable after the cessation of training, lasting at least on the order of one week. The perceptual learning demonstrated was shown to be elicited by both two-alternative forced choice and two-interval forced choice designs, with a strikingly similar time course and magnitude of narrowing. Moreover, the degree of narrowing was shown to be predictable on an individual basis, based upon the size of each participant's window as measured at baseline. The neural correlates of these changes, as assessed by event-related fMRI, were shown to be significant decreases in the BOLD signal during objectively synchronous and highly asynchronous conditions at loci in posterior superior temporal sulcus (pSTS), as well as in auditory and visual areas. These decreases were shown to have occurred selectively in trials wherein participants answered correctly, supporting increased efficiency of processing as a primary mechanism of the perceptual plasticity observed. Lastly, an increase in intrinsic functional coupling after training between pSTS and auditory areas and an increase in effective connectivity between pSTS and visual cortex during presentation of audiovisual stimulus pairs suggest a model in which pSTS becomes the

director of multimodal temporal processing improvements through enhanced coupling with unisensory visual and auditory areas. Taken as a whole, the evidence reported here is the first to describe fully the perceptual and neural changes accompanying a perceptual learning-induced narrowing of the multisensory temporal binding window.

Pertinent Themes

While the details of these results are discussed at length within their respective chapters, the results taken as a whole warrant discussion of certain over-arching themes that serve to unify them and form the foundations of further study. These themes are first discussed in the context of the present results and future directions pertaining to these issues are discussed at the end of this chapter.

The Role of Unisensory Temporal Processing and Bayesian Models of Integration

Multisensory integration has increasingly been viewed as the result of probabilistic processes that serve to combine information in a statistically optimal manner based upon the reliability of that information (for review, see (Knill and Pouget, 2004; Witten and Knudsen, 2005). Within this Bayesian framework, the brain takes the information available to it in the form of noisy sensory data and combines it with prior knowledge about the world to calculate an estimate, or posterior probability, that the world is in a particular state. By extension, posterior probabilities arising from bimodal information are thought to result from the combination of unimodal estimates, taking into account the variance of those probability distributions (Green and Swets, 1966; Witten and Knudsen, 2005).

This probabilistic framework has given rise to efforts to predict how information from the different senses might be combined in a wide range of contexts in a statistically optimal fashion. These concepts are illustrated in **Figure 4.1**, adapted from Witten and Knudsen (2005). In a typical experiment, participant responses in unimodal perceptual tasks (here, auditory and visual localization tasks) are first used to generate probability distributions (**Fig 4.1a**, red and blue curves). A prediction of the bimodal estimate (**Figure 4.1a**, black curve) is then generated as the product of these distributions and later compared with data from bimodal trials on the same perceptual task. In this way, Bayesian modeling has been used to accurately predict subjects' estimates of bimodal stimulus location in studies of the ventriloquist effect (Welch and Warren, 1980; Warren et al., 1981; Kording et al., 2007) and ventriloquist after-effect (Lewald, 2002). In these cases, the poor reliability of the auditory system in stimulus localization (seen as increased variance in participants' localization response distributions on auditory-alone trials, **Fig. 4.1b**) predicts a greater weighting of visual information (less variable by this measure) and a subsequent shift in bimodal stimulus localization toward the location of the visual stimulus. Thus, the ventriloquism effect is manifested as a shift in bimodal stimulus localization toward the location of the visual stimulus because the reliability of visual information is higher for localization tasks. This concept was tested recently in a study that manipulated visual stimulus reliability (with the use of visual stimuli in the form of Gaussian blobs of varying sizes) and measured its effects on bimodal stimulus localization

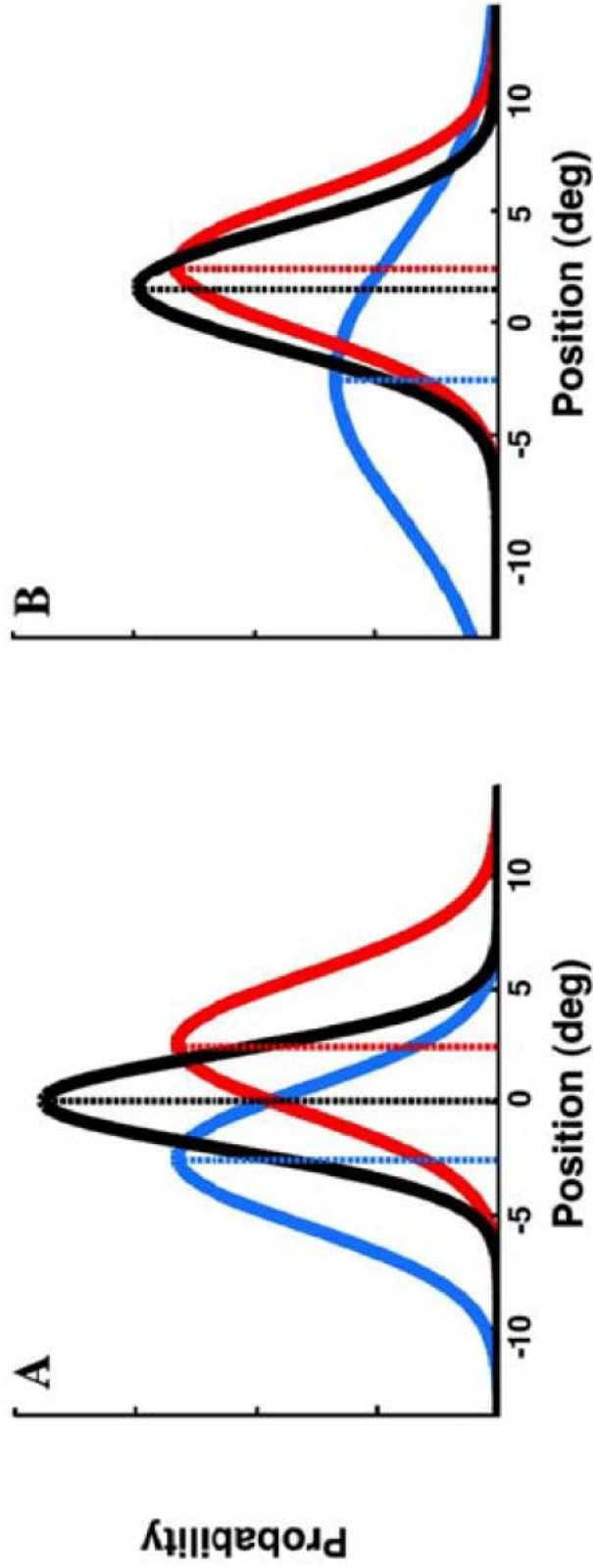


Figure 4.1. Bimodal localization estimates and the ventriloquist effect. *a.* Theoretical probability densities derived from auditory (blue) and visual (red) tasks, given equal reliability (variance) of the localization estimates for each. The audiovisual estimate is calculated as the product of these unimodal distributions (black). *b.* Realistic probability densities from auditory and visual localization estimates. In this case, auditory localization is far more variable than visual, and this is reflected in the localization estimate from bimodal trials, which is heavily biased toward the visual stimulus location. Adapted from Witten and Knudsen (2005).

(Alais and Burr, 2004). Results demonstrated that visually-biased responses dominated conditions in which visual stimulus reliability was high, but that a marked auditory bias emerged in trials containing unreliable visual stimuli.

These principles do not only govern the integration of multimodal spatial information: data derived from studies of visual-haptic size and angle discrimination (Hospedales and Vijayakumar, 2009), visual-vestibular integration in self-motion (Angelaki et al., 2009; Fetsch et al., 2009), comprehension of audiovisual speech in noise (Ma et al., 2009), and even multisensory enhancement in single neurons (Anastasio et al., 2000; Patton et al., 2002; Patton and Anastasio, 2003) have been placed within a Bayesian framework. More pertinently, the description of multisensory integration in temporal tasks as a Bayesian process has only recently been attempted. These studies have succeeded in predicting the pattern of integration in tasks centered on the sound-induced flash illusion (Shams et al., 2005) and audiovisual rate discrimination (Roach et al., 2006) based upon unimodal responses.

Extrapolating from these studies to the work at hand, it seems plausible that a similar framework could be used to model data derived from an audiovisual simultaneity judgment task. Specifically, such an analysis would produce an estimate of participants' temporal binding window as derived from estimates of unimodal temporal resolution and estimates of individual priors. A detailed proposal for such an approach is discussed in *Future Directions*, below.

Multisensory Brain Networks, Connectivity, and Oscillations

While the classic feed-forward model of multisensory integration has done much to explain how integration might occur through convergence of sensory inputs in multisensory association areas, recent work has pointed to brain regions previously thought to be the exclusive domain of individual senses (see Chapter I) as being important for multisensory interactions. However, the mechanisms by which integration might occur flexibly and specifically within an anatomically fixed distributed network are only now being delineated. One possible solution arises from the study of unisensory integration, which has indicated that correlated neural activity may be linked to feature binding (Eckhorn et al., 1988; Engel et al., 1992; Singer and Gray, 1995; Gray, 1999) and may even underlie the rapid cognitive switching required for online attentional modulation in individual sensory systems (Herrmann et al., 2004; Fries, 2005; Masuda, 2009). In this model, synchronization of neural signals allows specific patterns of functional connectivity to be established for the selection and integration of information within a distributed network (Salinas and Sejnowski, 2001; Fries, 2005; Womelsdorf et al., 2007). Alignment in the phases of these oscillations (phase coherence) has been shown to predict how strongly groups of interacting neurons may influence each other's response strength; as such, groups that are in phase with each other are likely to drive mutual responses, whereas groups that differ in phase are less likely—or sometimes, extremely unlikely—to influence each other (Womelsdorf et al., 2007). In a system that is distributed but nonetheless exhibits massive anatomical connections, manipulation of oscillatory phase is a fast, clean, and flexible way by which to control how nodes in this network might interact.

Not surprisingly, these concepts have recently been brought into the multisensory realm (see (Senkowski et al., 2008) for review) as a way by which binding of cross-modal information might occur in a distributed network of sensory areas. Since EEG was first used to study cross-modal binding (von Stein et al., 1999), finding that increased beta band coherence between parietal and temporal electrode sites correlated with correct semantic cross-modal matching, several studies have indicated that oscillations may be a critical component of multisensory binding. Studies of multisensory facilitation of reaction time (Senkowski et al., 2006), cross-modal sensory gating (Kisley and Cornwell, 2006), and simple concurrent presentation of audiovisual stimuli (Sakowitz et al., 2000; Sakowitz et al., 2001; Sakowitz et al., 2005) have since indicated that modulation of cortical oscillations may be critical in these processes. Additionally, the specific phenomenon of phase resetting—wherein the phases of oscillations in different cortical regions are brought into alignment after some event, theoretically making interaction between them more likely—has been described in cases of auditory-somatosensory and visual-auditory interactions (Lakatos et al., 2007; Kayser et al., 2008) and has been implicated as a major mechanism by which such interactions might occur. Finally, perception of a number of multisensory illusions including the sound induced flash illusion (Mishra et al., 2007; Kayser et al., 2008; Mishra et al., 2008), the McGurk effect (Kaiser et al., 2005; Kaiser et al., 2007), and the rubber hand illusion (Kanayama et al., 2007, 2009) has been shown to correlate with the power of evoked beta- and gamma-band activity. These interactions have been hypothesized by some to represent a signature of audiovisual binding (Senkowski et al., 2008). In fact, a very recent modeling study investigating how some of the Bayesian processes described in the previous section

might be instantiated on a neural level has settled on neuronal oscillations as the most likely mechanism for cross-modal binding in distributed neural networks (Bobrowski et al., 2009).

Understanding multisensory interactions as phenomena possibly arising from a distributed network of unisensory and multisensory areas is critical for interpretation of the results described in this volume. In fact, the dynamic causal models proposed in Chapter III explicitly reflect the reciprocal statistical dependence of pSTS and unisensory visual and auditory regions (Stephan et al., 2007; Kasess et al., 2009; Stephan et al., 2009). The modeling results reported here suggest an increased influence of activity in pSTS on activity in visual cortex but do not directly identify the manner by which such an influence might be instantiated. Given that phase resetting has been robustly demonstrated in studies of cross-modal influence (Lakatos et al., 2007; Kayser et al., 2008), this seems to be a particularly promising mechanism by which pSTS could influence visual cortical activity. Thus, the increased influence of pSTS on visual areas observed after training could reflect more effective phase locking of oscillatory activity between the two areas during simultaneous audiovisual presentation, thereby resulting in more efficient binding. This fits very well with the fact that the signature of perceptual learning is itself a BOLD signal decrease, implying more efficient processing in these regions after training.

While it has succeeded in identifying a network whose dynamics change after training, functional imaging simply does not have sufficient temporal resolution to reveal detailed changes in neural synchronization. Thus, a critical next step must be the utilization of electrophysiology recorded at the scalp (i.e., EEG, ERP) in order to identify

whether changes in neuronal oscillations may accompany the plastic changes described here. A proposal for such a study and predictions as to what might be observed are found in *Future Directions*, below.

Clinical Implications

One of the rationales for studying the plasticity of the multisensory temporal binding window is that multisensory temporal processing appears to be altered in certain clinical populations. These deficits are thoroughly reviewed in Chapters II and III, but because the processes described here appear to depend critically upon binding of information across nodes in a distributed network, because emerging evidence suggests that deficits in information binding and the oscillatory activity often associated with it may be organizing themes in several of these disorders (Uhlhaas and Singer, 2006), and because knowledge of these deficits may allow us to make specific predictions about multisensory function and plasticity in these clinical populations, such deficits are briefly reviewed below.

Evidence that pan-sensory and specifically multisensory deficits exist in developmental dyslexia has accumulated over the past decade, punctuated by findings of an enlarged multisensory temporal binding window in these individuals (Hairston et al., 2005) as well as abnormal BOLD measures in pSTS during visual-auditory linguistic tasks (Rumsey et al., 1992; Brunswick et al., 1999; Paulesu et al., 2001; Shaywitz et al., 2002; Blau et al., 2009; Blau et al., 2010) and abnormal functioning in multisensory cerebellar areas (Fawcett et al., 2001; Stoodley et al., 2005; Stoodley and Stein, 2009). Additionally, in the realm of oscillations and coherence, individuals with dyslexia clearly

exhibit abnormal (but often conflicting) patterns. Early ERP studies of long-range functional connectivity during word-reading tasks in dyslexic children revealed decreased inter-hemispheric beta-band activity but increased local gamma-band activity, indicating a specific deficit in binding among nodes of a distributed network (Sklar et al., 1972; Leisman and Ashkenazi, 1980). More recent work in MEG and ERP showed decreased coherence in the beta- and gamma-band ranges in auditory regions of adults with dyslexia when compared with normal controls (Nagarajan et al., 1999; Ucles et al., 2009). Interestingly, a reversal of these deficits appears during rest (Marosi et al., 1995; Shiota et al., 2000), indicating that deficits in long-range binding may be task-specific. Taken together, this evidence indicates that developmental dyslexia is characterized by a task-specific deficit in binding audiovisual information across disparate brain regions. Because the training paradigm described here appears capable of effecting improvements in precisely these processes, it may form the basis of promising remediation paradigms in the future.

In addition to the deficits in multisensory processing (Ross et al., 2007; de Jong et al., 2009; Jardri et al., 2009) and pSTS function (Szyck et al., 2009), there is a robust literature on oscillatory and feature binding deficits in individuals with schizophrenia. In fact, current theories of the pathophysiology of schizophrenia have focused upon deficits in coordination of distributed processes across multiple cortical areas that may produce some of the core cognitive deficits that are characteristic of the disorder (Dolan et al., 1999; Friston, 1999; Phillips and Silverstein, 2003). Consistent with these theories, individuals with schizophrenia exhibit significant impairments in perceptual grouping (Uhlhaas and Silverstein, 2005), and unisensory tasks assessing visual binding and

backward masking (Spencer et al., 2003; Spencer et al., 2004; Wynn et al., 2005), as well as detection of auditory oddball stimuli (Gallinat et al., 2004), consistently reveal decreases in performance and in evoked beta- and gamma-band activity relative to controls. Like those observed in dyslexia, the connectivity differences observed in schizophrenia appear to be focused on long-distance projections: studies of induced activity during a facial recognition task revealed decreased large-scale but increased local synchronization (Uhlhaas et al., 2006b; Uhlhaas et al., 2006a); cross-cortical measures of functional connectivity in fMRI consistently reveal decreases in individuals with schizophrenia on a wide range of cognitive and perceptual tasks (Schlosser et al., 2003; Honey et al., 2005; Meyer-Lindenberg et al., 2005); and several anatomical studies using diffusion tensor imaging have revealed large decreases in long-distance cortico-cortical projections but increases in local white matter (Hubl et al., 2004; Kubicki et al., 2005). Lastly, abnormalities in the cortical GABAergic interneurons (Lewis et al., 2005) and NMDA receptors (Moghaddam, 2003; Moghaddam and Jackson, 2003) that are thought to be crucial for synchronization of neuronal activity have been robustly demonstrated in schizophrenia. Overall, individuals with schizophrenia exhibit perceptual binding deficits in the context of impaired neural synchronization, a combination that could also result in abnormal multisensory temporal function given the current results. Thus far, however, no such deficits have yet been demonstrated, and investigation of multisensory temporal function in schizophrenia must be considered as a potentially fruitful area of future inquiry.

Lastly, efforts by our laboratory (Foss-Feig et al., 2010) have recently identified specific dysfunctions in multisensory temporal function in autism. This fits very well

with the concept of *weak central coherence*, which has become a leading theory of autism pathophysiology (see (Happé and Frith, 2006) for review). Positing that the core deficit in autism may be defined as a failure to bind feature information into a holistic percept, this theory has recently drawn links between the deficits observed in this realm and abnormal patterns of neuronal synchronization seen in autism. In several fMRI studies, decreased functional connectivity was observed among frontal, parietal, and superior temporal lobes on a wide variety of linguistic and cognitive tasks (Castelli et al., 2002; Just et al., 2004; Kana et al., 2006; Koshino et al., 2008); evidence in ERP of decreased gamma-band activity during facial recognition tasks and passive presentation of auditory stimuli has recently been described (Grice et al., 2001; Wilson et al., 2007); finally, developmental anatomical abnormalities in autism consisting of transient increases in white matter followed by anatomical hypoconnectivity point to a possible anatomical substrate for the abnormalities seen in functional connectivity (Herbert et al., 2003; Just et al., 2004; Kennedy and Courchesne, 2008). Like those seen in dyslexia, the combination of deficits seen in autism point to abnormal multisensory processing in the context of abnormal functional connectivity between distributed networks of sensory and higher-level cortical areas. As in dyslexia and schizophrenia, determination of the flexibility of these deficits in the face of multisensory perceptual learning is a potentially fruitful avenue for future clinically-oriented research. Predictions pertaining to the size of the multisensory temporal binding window, its flexibility, and its relationship to the clinical manifestations of these disorders may be found in *Future Directions* below.

Future Directions

Drawing on the work described here and the context provided above, it is possible to identify several studies that would significantly expand our current knowledge of multisensory temporal processing, multisensory plasticity, and the networks that underlie these phenomena.

As alluded to in the above section on a Bayesian framework for multisensory integration, it should be possible to predict the size and shape of individuals' multisensory temporal binding windows if given measures of their unisensory temporal discrimination abilities and some predictor of their priors. Doing so would represent the first attempt to draw explicit links between individual unisensory temporal processing and the likelihood of perceiving audiovisual simultaneity and would provide a framework for describing the changes brought about by perceptual training in probabilistic terms.

The approach to creating such a model might involve estimation of the appropriate likelihood ratio and priors based upon individual performance data and classical Bayesian formulations. Theoretically, this approach may resemble the following. Given the basic assumptions of signal detection theory, a simultaneity judgment may be described as the generation of an estimate of probability that the onsets of auditory and visual events are synchronous given the timing of noisy neural signals produced in response to said events. Thus, given a quantity T corresponding to the difference in timing between neural auditory and visual events a and v , a customization of Bayes' rule gives:

$$P(A = V | T) = \frac{P(T | A = V)}{P(T)} P(A = V) \quad , \quad (1)$$

where A and V represent the onset times of auditory and visual events in the outside world. Because the probability that T could occur is invariant across stimulus space, it may be described as a constant (k), leading to a simplification of Equation 1:

$$P(A = V | T) = P(T | A = V) * kP(A = V). \quad (2)$$

Thus, given an estimate of the likelihood that a neural time difference T could be generated by truly synchronous auditory and visual events, as well as an estimate of the prior likelihood that event times A and V are the same (the Bayesian prior; right-most term), one may derive an estimate of the likelihood that these auditory and visual events occurred synchronously.

In order to determine the likelihood that T could be produced by the synchronous occurrence of auditory and visual events, estimates of the latencies and spreads of unisensory neural responses must be obtained. As per the approach recently adopted by Colonius and Diederich for a similar derivation (Colonius et al., 2010), simple auditory and visual reaction time tasks might be used to obtain said estimates. In this way, cumulative probability functions may be derived from response time data (**Fig. 4.2a**) and differentiated to produce Gaussian probability density functions (**Fig. 4.2b**) whose spreads correspond to the participant's variability in reaction time to the auditory or visual stimulus and whose latency corresponds to the time to peak of the response after stimulus onset. Because both the auditory and visual distributions are locked to stimulus onset, the difference in time between the peaks of these functions may serve as an

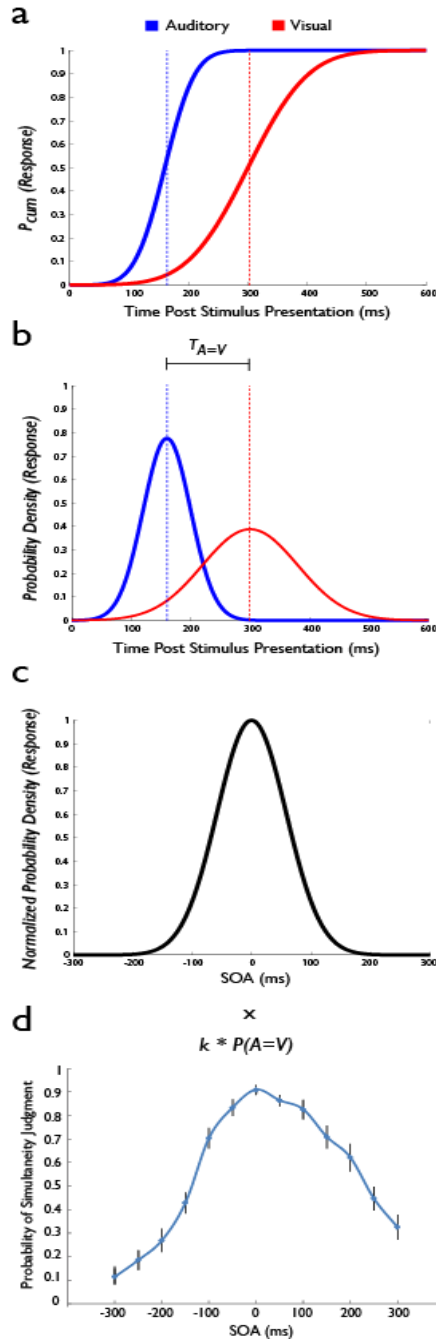


Figure 4.2. *A proposed approach to modeling probability of audiovisual simultaneity judgment based upon unisensory temporal data. a) Cumulative probability functions derived from data obtained in auditory and visual reaction time tasks. b) Differentiation of the functions shown in a results in probability density functions representing peak response latency and variance of response to visual and auditory stimuli. c) The product of these distributions provides a Gaussian function in which the probability that neural peak time difference T was produced by synchronous events varies by SOA. d) Linear transformation of this function with non-uniform priors may give an estimate of the probability of simultaneity judgment across SOAs in individual participants.*

estimate of the quantity T when event presentation times A and V are the same. It should be noted that this formulation deems the estimated latency difference $T_{A=V}$ to be due to a difference in perceptual latency, with roughly equal motor processing times between the two modalities. The product of these distributions may be calculated to give a Gaussian function centered on $T_{A=V}$ giving the probability based upon T that the neural responses could have been produced by synchronous audiovisual events (**Fig. 4.2c**), which is the value of the left-most term in Equation 2 above.

The Bayesian prior (the right-most term in Equation 2) may reflect the observer's experience, hard-wired biases imposed by the neuroanatomical and neurophysiological states, or even biases imposed by the task (Shams et al., 2005). Although priors are often deemed to be uniform across the sample space, thereby conferring equal probability to all possible outcomes, a non-uniform prior would be appropriate in this framework, reflecting prior knowledge that pairs on the right side of the distribution (wherein visual precedes auditory) are more likely to have been caused by the same external event than those on the left side of the distribution. This model predicts that participants are more likely to judge audiovisual simultaneity in visual-before-auditory conditions, a pattern that fits the data acquired by our laboratory (Chapter II) as well as others (Zampini et al., 2005b; Zampini et al., 2005a).

Although priors are seldom measured directly in traditional Bayesian analyses, doing so in this case would provide for a more complete model of simultaneity perception. A solution to this problem was recently found by Shams and colleagues (Shams et al., 2005) for the estimation of priors given unisensory and multisensory responses during a task based upon the sound induced flash illusion. Given prior

estimates derived from one pool of participants, they successfully predicted the multisensory responses of a second pool given their auditory- and visual-alone responses. A similar solution may be attempted here. In the case of audiovisual simultaneity judgments, the unknown priors (scaled by the constant k) may be taken to be the quotient of the observed probability of simultaneity judgment and the predicted probability given by the likelihood ratio at a given SOA. Given the circularity of this estimate (it is based upon an inversion of Bayes' rule), it could not be used for prediction of a multisensory simultaneity distribution given only one data set, but could be used for estimation of averaged population-based priors and for subsequent prediction of multisensory distributions in a second data set (**Fig. 4.2d**).

Within this framework, decreases in probability of simultaneity judgment at non-zero SOAs after training could be brought about by decreases in prior likelihood of simultaneity at those SOAs or an increase in the temporal reliability of unisensory neural responses themselves. Given the data obtained thus far, a change in priors is supported by the asymmetrical window narrowing observed in Chapter II; indeed, feedback has been shown to be critical for altering experience-based priors (Backus and Haijiang, 2007; Di Luca et al., 2009). However, the idea that temporal window narrowing may be the result of improvements in unimodal temporal reliability is supported by data reflecting an increase in effective connectivity between multisensory pSTS and visual areas (Chapter III) as well as by evidence that visual gap detection capabilities may be enhanced after training (Appendix). It is likely that both mechanisms may be at play, and carefully-designed experiments meant to assess unisensory temporal processing before

and after audiovisual simultaneity judgment training may help to differentiate between them.

A more complete picture of the changes wrought by multisensory perceptual learning could also be derived with the use of more temporally-sensitive methods for detecting changes in neural function on the network level. As mentioned above, the increased influence of pSTS on visual areas observed after training could reflect more complete phase locking of oscillatory activity between the two regions during simultaneous audiovisual presentation. However, this hypothesis is based upon indirect estimations of neural activity. A better approach to assess the connectivity changes implied by the data in Chapter III would be to utilize the temporal resolution offered by direct measures of neural activity such as ERP and MEG. Several predictions may be generated should such electrophysiological data be acquired before and after training on the audiovisual simultaneity judgment task. First, based upon our own data and previous studies of effective connectivity during simultaneity perception (Noesselt et al., 2007), we may hypothesize that evoked beta- and/or gamma-band activity at temporal and occipital electrodes will vary by SOA based upon the probability of simultaneity judgment for each individual at that SOA. Second, a shift in phase coherence in these bands at these electrode sites should reflect the shifts in probability of simultaneity judgment observed after training; thus, simultaneity perception at truly synchronous trials should result in more precise synchronization of oscillations, whereas a drop in simultaneity perception at truly asynchronous SOAs should result in these oscillations being more out of phase. Lastly, the magnitude of these shifts should correlate with the degree of temporal window narrowing observed behaviorally. Overall, using more direct

and temporally precise methods of measuring neural oscillatory activity will allow for a clearer picture of how the nodes in the sensory network described here might interact during simultaneity perception and how they might change after multisensory perceptual training. Importantly, the information gathered in such a study would be temporally precise enough to allow for definition of set epochs of post-stimulus time wherein interactions are thought to occur. Knowledge of these time periods would be extremely important in the development of a subsequent TMS study meant to disrupt multisensory perceptual learning by temporarily creating a functional lesion in pSTS.

From a clinical perspective, the perceptual learning paradigm described in this work represents an unprecedented means by which circuits thought to be abnormal in dyslexia, autism, and possibly schizophrenia might be intentionally altered. It is a logical goal, then, that both assessment of multisensory temporal binding window size and training on this paradigm should be attempted in these clinical populations. While it is difficult to predict the exact outcomes of these efforts, several bits of evidence from the literature may predict variable rates of success. First, while an enlarged multisensory temporal binding window has been demonstrated in both dyslexia and autism on different tasks, there is not nearly as much data to suggest altered multisensory temporal processing in schizophrenia. Thus, while an enlarged temporal binding window is expected to be observed in individuals with dyslexia and autism as measured via the audiovisual simultaneity judgment task, it is more difficult to make a specific prediction as to the size of that window in participants suffering from schizophrenia. On a methodological note, it should be stated that task demands in these clinical populations in particular must be considered carefully before any data collection is attempted in order to

ensure interpretability of results. For example, it is known that working memory function is impaired in schizophrenia and that dyslexia carries with it attention deficit disorder as a common comorbidity. Modification of the currently used 2-IFC paradigm to a cross-modal audiovisual temporal order judgment might allow future researchers to avoid working memory confounds in the schizophrenic population, and a more efficient route to left and right window sizes (i.e., through the use of adaptive staircase procedures) may help minimize the effects of attentional issues in the dyslexic population in the future. If and when window size measurements are derived for these individuals, and if an effect is observed, these measurements could provide great insight into the relationship between multisensory temporal processing and the specific symptoms of these disorders.

Second, assessing the ability of the multisensory temporal binding window to be narrowed in each of these populations will be extremely informative. As mentioned above, the types of oscillations thought to underlie multisensory binding appear to be fundamentally altered in all three disorders, and abnormalities in neurotransmitter function have also been hypothesized to be the cause of these alterations in all three disorders. If the cellular machinery used to regulate neural oscillations is fundamentally disordered and if controlled alteration of these oscillations is the basis of the narrowing observed in the multisensory temporal binding window, one might predict that such a narrowing might not be possible in these clinical populations. Regardless of the outcome of this assay, however, the results would doubtless be informative as to the underlying basis of disease in individuals with these disorders.

Lastly, if accurate measurement of the multisensory temporal binding window is possible in individuals with dyslexia, autism and schizophrenia, if that window is

demonstrated to be abnormal in these individuals, and if that window can be narrowed with perceptual training, future researchers may be presented with a rare opportunity to directly assess the likelihood of a causal relationship between abnormal multisensory temporal processing and the clinical symptoms of these disorders. Caution should be urged in the interpretation of even these results, however, as the alteration of abnormal multisensory circuits that may have caused disruption during development—for example, while learning to read—may have no effect on clinical measures regardless of its causal role in the etiology of the disorder being studied. In this case, study of training-induced window size narrowing in a developing population may be the optimal approach.

Despite the difficulties inherent in these translational efforts, it should be emphasized that one of the most important possible contributions of the work described in this volume may be in the development of improved diagnostic measures and novel remediation approaches to aid those who are afflicted with disorders of multisensory processing.

Conclusions

The work described in this volume represents the first attempt to define and alter the size of the multisensory temporal binding window with the use of an audiovisual simultaneity judgment paradigm. Our results demonstrate that the resultant narrowing in the size of the multisensory temporal binding window is rapid, robust, and stable across time. Moreover, we have shown that decreases in BOLD signal during presentation of synchronous and highly asynchronous audiovisual stimulus pairs in multisensory and unisensory regions accompany the behavioral changes induced by training, and that

selective coupling among these regions may underlie the perception of audiovisual simultaneity. These results form the foundation of future studies wherein the probabilistic and neural bases of these processes might be described in more detail. Finally, it is our hope that this first demonstration of multisensory plasticity in normative adults may be the beginning of a process that leads to the more effective diagnosis and treatment of multisensory processing disorders.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257-262.

Anastasio TJ, Patton PE, Belkacem-Boussaid K (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Comput* 12:1165-1187.

Angelaki DE, Klier EM, Snyder LH (2009) A vestibular sensation: probabilistic approaches to spatial perception. *Neuron* 64:448-461.

Backus BT, Haijiang Q (2007) Competition between newly recruited and pre-existing visual cues during the construction of visual appearance. *Vision Res* 47:919-924.

Blau V, van Atteveldt N, Ekkebus M, Goebel R, Blomert L (2009) Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr Biol* 19:503-508.

Blau V, Reithler J, van Atteveldt N, Seitz J, Gerretsen P, Goebel R, Blomert L (2010) Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133:868-879.

Bobrowski O, Meir R, Eldar YC (2009) Bayesian filtering in spiking neural networks: noise, adaptation, and multisensory integration. *Neural Comput* 21:1277-1320.

Brunswick N, McCrory E, Price CJ, Frith CD, Frith U (1999) Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: A search for Wernicke's Wortschatz? *Brain* 122 (Pt 10):1901-1917.

Castelli F, Frith C, Happe F, Frith U (2002) Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. In, p 1839: Oxford Univ Press.

Colonius H, Diederich A, Colonius H (2010) The optimal time window of visual-auditory integration: a reaction time analysis. In: *Front. Integr. Neurosci.*

de Jong JJ, Hodiament PP, Van den Stock J, de Gelder B (2009) Audiovisual emotion recognition in schizophrenia: reduced integration of facial and vocal affect. *Schizophr Res* 107:286-293.

Di Luca M, Machulla TK, Ernst MO (2009) Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. *J Vis* 9:7 1-16.

Dolan RJ, Fletcher PC, McKenna P, Friston KJ, Frith CD (1999) Abnormal neural integration related to cognition in schizophrenia. *Acta Psychiatr Scand Suppl* 395:58-67.

Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1988) Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol Cybern* 60:121-130.

Engel AK, Konig P, Kreiter AK, Schillen TB, Singer W (1992) Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci* 15:218-226.

Fawcett AJ, Nicolson RI, Maclagan F (2001) Cerebellar tests differentiate between groups of poor readers with and without IQ discrepancy. *J Learn Disabil* 34:119-135.

Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE (2009) Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci* 29:15601-15612.

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2010) An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res* 203:381-389.

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9:474-480.

Friston KJ (1999) Schizophrenia and the disconnection hypothesis. *Acta Psychiatr Scand Suppl* 395:68-79.

Gallinat J, Winterer G, Herrmann CS, Senkowski D (2004) Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing. *Clin Neurophysiol* 115:1863-1874.

Gray CM (1999) The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron* 24:31-47, 111-125.

Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. New York,: Wiley.

Grice SJ, Spratling MW, Karmiloff-Smith A, Halit H, Csibra G, de Haan M, Johnson MH (2001) Disordered visual processing and oscillatory brain activity in autism and Williams syndrome. In, p 2697.

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT (2005) Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res* 166:474-480.

Happé F, Frith U (2006) The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J Autism Dev Disord* 36:5-25.

Herbert MR, Ziegler DA, Deutsch CK, O'Brien LM, Lange N, Bakardjiev A, Hodgson J, Adrien KT, Steele S, Makris N (2003) Dissociations of cerebral cortex, subcortical and cerebral white matter volumes in autistic boys. In, p 1182: Oxford Univ Press.

Herrmann CS, Munk MH, Engel AK (2004) Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn Sci* 8:347-355.

Honey GD, Pomarol-Clotet E, Corlett PR, Honey RA, McKenna PJ, Bullmore ET, Fletcher PC (2005) Functional dysconnectivity in schizophrenia associated with attentional modulation of motor function. *Brain* 128:2597-2611.

Hospedales T, Vijayakumar S (2009) Multisensory oddity detection as bayesian inference. *PLoS One* 4:e4205.

Hubl D, Koenig T, Strik W, Federspiel A, Kreis R, Boesch C, Maier SE, Schroth G, Lovblad K, Dierks T (2004) Pathways that make voices: white matter changes in auditory hallucinations. *Arch Gen Psychiatry* 61:658-668.

Jardri R, Pins D, Bubrowszky M, Lucas B, Lethuc V, Delmaire C, Vantuyghem V, Desprez P, Thomas P (2009) Neural functional organization of hallucinations in schizophrenia: multisensory dissolution of pathological emergence in consciousness. *Conscious Cogn* 18:449-457.

Just MA, Cherkassky VL, Keller TA, Minshew NJ (2004) Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. In, p 1811: Oxford Univ Press.

Kaiser J, Lennert T, Lutzenberger W (2007) Dynamics of oscillatory activity during auditory decision making. *Cereb Cortex* 17:2258-2267.

Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing lips: gamma-band activity during audiovisual speech perception. *Cereb Cortex* 15:646-653.

Kana RK, Keller TA, Cherkassky VL, Minshew NJ, Just MA (2006) Sentence comprehension in autism: thinking in pictures with decreased functional connectivity. In, p 2484: Oxford Univ Press.

Kanayama N, Sato A, Ohira H (2007) Crossmodal effect with rubber hand illusion and gamma-band activity. *Psychophysiology* 44:392-402.

Kanayama N, Sato A, Ohira H (2009) The role of gamma band oscillations and synchrony on rubber hand illusion and crossmodal integration. *Brain Cogn* 69:19-29.

Kasess CH, Stephan KE, Weissenbacher A, Pezawas L, Moser E, Windischberger C (2009) Multi-subject analyses with dynamic causal modeling. *Neuroimage* 49:3065-3074.

Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18:1560-1574.

Kennedy DP, Courchesne E (2008) The intrinsic functional organization of the brain is altered in autism. In, pp 1877-1885: Elsevier.

Kisley MA, Cornwell ZM (2006) Gamma and beta neural activity evoked during a sensory gating paradigm: effects of auditory, somatosensory and cross-modal stimulation. *Clin Neurophysiol* 117:2549-2563.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712-719.

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. *PLoS One* 2:e943.

Koshino H, Kana RK, Keller TA, Cherkassky VL, Minshew NJ, Just MA (2008) fMRI investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. In, p 289: Oxford Univ Press.

Kubicki M, McCarley RW, Shenton ME (2005) Evidence for white matter abnormalities in schizophrenia. *Curr Opin Psychiatry* 18:121-134.

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-292.

Leisman G, Ashkenazi M (1980) Aetiological factors in dyslexia: IV. Cerebral hemispheres are functionally equivalent. *Int J Neurosci* 11:157-164.

Lewald J (2002) Rapid adaptation to auditory-visual spatial disparity. *Learn Mem* 9:268-278.

Lewis DA, Hashimoto T, Volk DW (2005) Cortical inhibitory neurons and schizophrenia. *Nat Rev Neurosci* 6:312-324.

Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC (2009) Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One* 4:e4638.

Marosi E, Harmony T, Becker J, Reyes A, Bernal J, Fernandez T, Rodriguez M, Silva J, Guerrero V (1995) Electroencephalographic coherences discriminate between children with different pedagogical evaluation. *Int J Psychophysiol* 19:23-32.

Masuda N (2009) Selective population rate coding: a possible computational role of gamma oscillations in selective attention. *Neural Comput* 21:3335-3362.

Meyer-Lindenberg AS, Olsen RK, Kohn PD, Brown T, Egan MF, Weinberger DR, Berman KF (2005) Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia. *Arch Gen Psychiatry* 62:379-386.

Mishra J, Martinez A, Hillyard SA (2008) Cortical processes underlying sound-induced flash fusion. *Brain Res* 1242:102-115.

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci* 27:4120-4131.

Moghaddam B (2003) Bringing order to the glutamate chaos in schizophrenia. *Neuron* 40:881-884.

Moghaddam B, Jackson ME (2003) Glutamatergic animal models of schizophrenia. *Ann N Y Acad Sci* 1003:131-137.

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM (1999) Cortical auditory signal processing in poor readers. *Proc Natl Acad Sci U S A* 96:6483-6488.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431-11441.

Patton P, Belkacem-Boussaid K, Anastasio TJ (2002) Multimodality in the superior colliculus: an information theoretic analysis. *Brain Res Cogn Brain Res* 14:10-19.

- Patton PE, Anastasio TJ (2003) Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons. *Neural Comput* 15:783-810.
- Paulesu E, Demonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD, Frith U (2001) Dyslexia: cultural diversity and biological unity. *Science* 291:2165-2167.
- Phillips WA, Silverstein SM (2003) Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. *Behav Brain Sci* 26:65-82; discussion 82-137.
- Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc Biol Sci* 273:2159-2168.
- Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ (2007) Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr Res* 97:173-183.
- Rumsey JM, Andreason P, Zametkin AJ, Aquino T, King AC, Hamburger SD, Pikus A, Rapoport JL, Cohen RM (1992) Failure to activate the left temporoparietal cortex in dyslexia. An oxygen 15 positron emission tomographic study. *Arch Neurol* 49:527-534.
- Sakowitz OW, Schurmann M, Basar E (2000) Oscillatory frontal theta responses are increased upon bisensory stimulation. *Clin Neurophysiol* 111:884-893.
- Sakowitz OW, Quiroga RQ, Schurmann M, Basar E (2001) Bisensory stimulation increases gamma-responses over multiple cortical regions. *Brain Res Cogn Brain Res* 11:267-279.
- Sakowitz OW, Quiroga R, Schurmann M, Basar E (2005) Spatio-temporal frequency characteristics of intersensory components in audiovisually evoked potentials. *Brain Res Cogn Brain Res* 23:316-326.
- Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nat Rev Neurosci* 2:539-550.

Schlosser R, Gesierich T, Kaufmann B, Vucurevic G, Hunsche S, Gawehn J, Stoeter P (2003) Altered effective connectivity during working memory performance in schizophrenia: a study with fMRI and structural equation modeling. *Neuroimage* 19:751-763.

Senkowski D, Molholm S, Gomez-Ramirez M, Foxe JJ (2006) Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: a high-density electrical mapping study. *Cereb Cortex* 16:1556-1565.

Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401-409.

Shams L, Ma WJ, Beierholm U (2005) Sound-induced flash illusion as an optimal percept. *Neuroreport* 16:1923-1927.

Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2002) Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol Psychiatry* 52:101-110.

Shiota M, Koeda T, Takeshita K (2000) Cognitive and neurophysiological evaluation of Japanese dyslexia. *Brain Dev* 22:421-426.

Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18:555-586.

Sklar B, Hanley J, Simmons WW (1972) An EEG experiment aimed toward identifying dyslexic children. *Nature* 240:414-416.

Spencer KM, Nestor PG, Niznikiewicz MA, Salisbury DF, Shenton ME, McCarley RW (2003) Abnormal neural synchrony in schizophrenia. *J Neurosci* 23:7407-7411.

Spencer KM, Nestor PG, Perlmuter R, Niznikiewicz MA, Klump MC, Frumin M, Shenton ME, McCarley RW (2004) Neural synchrony indexes disordered perception and cognition in schizophrenia. *Proc Natl Acad Sci U S A* 101:17288-17293.

Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38:387-401.

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ (2009) Ten simple rules for dynamic causal modeling. *Neuroimage* 49:3099-3109.

Stoodley CJ, Stein JF (2009) The cerebellum and dyslexia. *Cortex*.

Stoodley CJ, Fawcett AJ, Nicolson RI, Stein JF (2005) Impaired balancing ability in dyslexic children. *Exp Brain Res* 167:370-380.

Szyck GR, Munte TF, Dillo W, Mohammadi B, Samii A, Emrich HM, Dietrich DE (2009) Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. *Schizophr Res* 110:111-118.

Ucles P, Mendez M, Garay J (2009) Low-level defective processing of non-verbal sounds in dyslexic children. *Dyslexia* 15:72-85.

Uhlhaas PJ, Silverstein SM (2005) Perceptual organization in schizophrenia spectrum disorders: empirical research and theoretical implications. *Psychol Bull* 131:618-632.

Uhlhaas PJ, Singer W (2006) Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron* 52:155-168.

Uhlhaas PJ, Phillips WA, Mitchell G, Silverstein SM (2006a) Perceptual grouping in disorganized schizophrenia. *Psychiatry Res* 145:105-117.

Uhlhaas PJ, Linden DE, Singer W, Haenschel C, Lindner M, Maurer K, Rodriguez E (2006b) Dysfunctional long-range coordination of neural activity during Gestalt perception in schizophrenia. *J Neurosci* 26:8168-8175.

von Stein A, Rappelsberger P, Sarnthein J, Petsche H (1999) Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cereb Cortex* 9:137-150.

Warren DH, Welch RB, McCarthy TJ (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept Psychophys* 30:557-564.

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638-667.

Wilson TW, Rojas DC, Reite ML, Teale PD, Rogers SJ (2007) Children and adolescents with autism exhibit reduced MEG steady-state gamma responses. In, pp 192-197: Elsevier.

Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. *Neuron* 48:489-496.

Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P (2007) Modulation of neuronal interactions through neuronal synchronization. *Science* 316:1609-1612.

Wynn JK, Light GA, Breitmeyer B, Nuechterlein KH, Green MF (2005) Event-related gamma activity in schizophrenia patients during a visual backward-masking task. *Am J Psychiatry* 162:2330-2336.

Zampini M, Shore DI, Spence C (2005a) Audiovisual prior entry. *Neurosci Lett* 381:217-222.

Zampini M, Guest S, Shore DI, Spence C (2005b) Audio-visual simultaneity judgments. *Percept Psychophys* 67:531-544.

APPENDIX

PERCEPTUAL TRAINING ON AN AUDIOVISUAL SIMULTANEITY JUDGMENT TASK ALTERS SOUND-INDUCED FLASH ILLUSION (SIFI) PERFORMANCE

Abstract

Life in a multisensory world requires the accurate integration of stimuli from the different senses. In this process, the temporal relationship between sensory stimuli has been shown to be critical in determining which of these share a common environmental origin. The literature has increasingly indicated that there is a window in time within which audiovisual stimuli are likely to be perceptually bound, here referred to as the multisensory temporal binding window. While this window's boundaries have been delineated time and again in psychophysical and neurophysiological investigations, its malleability has only recently been demonstrated. Specifically, a recent publication by this group has indicated that perceptual training on an audiovisual simultaneity judgment paradigm is capable of eliciting a robust and stable narrowing of this window. Here we investigated whether this narrowing may have been accompanied by changes in performance on a task based upon the sound-induced flash illusion (SIFI). A subset of training participants underwent assessment on the SIFI task both before and after training. Results demonstrated an increase in sensitivity in training but not control (exposure) participants on the SIFI task, an effect that scaled with the degree of temporal window narrowing exhibited by individual participants. Intriguingly, this increase in sensitivity was driven not by a decrease in propensity to report illusory flashes, but by the

correct identification of rapidly-presented two-flash conditions. These results indicate that gains in multisensory temporal precision brought about by perceptual training are capable of transfer to visual temporal processing.

Introduction

Our ability to perceive the world in an accurate and meaningful way depends critically upon the appropriate integration of cross-modal stimuli. One of the more difficult aspects of this process comes in determining which of a constant stream of stimuli from the different senses were caused by the same environmental event. One of the strategies the brain employs in accomplishing this task relies upon the temporal structure of multisensory stimulus pairs; those events that occur in close temporal proximity are likely to have been caused by the same environmental event, whereas events that occur at extremely disparate times are unlikely to have a common origin. However, because the environmental energies carrying this information (i.e., light, sound) propagate at very different rates, this relationship must be flexibly specified. Thus, the construct of a temporal window of multisensory binding has been proposed, within which paired cross-modal events are likely to produce enhanced neural, perceptual, and behavioral responses (Dixon and Spitz, 1980; McGrath and Summerfield, 1985; Pandey et al., 1986; Meredith et al., 1987; Colonius and Diederich, 2004; Diederich and Colonius, 2004, 2008; Colonius et al., 2009; Diederich and Colonius, 2009). Most studies of multisensory temporal processing to date have focused upon the realm of audiovisual interactions, capitalizing upon such tasks as audiovisual

simultaneity judgment to determine the boundaries of the multisensory temporal binding window.

Studies have highlighted the developmental dynamics of the multisensory temporal binding window (Lewkowicz, 1996), but its ability to be narrowed with perceptual training has only recently been uncovered (Powers et al., 2009). Whether the improvements in multisensory temporal acuity wrought by training are capable of affecting processing in single modalities or in other multisensory tasks has yet to be determined, however. Reports of generalization of effects are rare in the perceptual learning literature (Doshier and Lu, 2007; Roth et al., 2008; Jeter et al., 2009; Lapid et al., 2009; Polat, 2009), and while improvements in auditory temporal processing following training on a somatosensory temporal discrimination task (Nagarajan et al., 1998), no study has shown transfer of learning from a multisensory task to one involving only single sensory modalities. We engaged a subset of multisensory perceptual training and exposure control participants in a task based upon the sound-induced flash illusion (SIFI; (Shams et al., 2000, 2002) both before and after training in order to determine whether successful narrowing of the multisensory temporal binding window might alter their performance on this different but related multisensory temporal task.

Materials And Methods

2-AFC Training

Subjects

Twelve (12) Vanderbilt undergraduate and graduate students (mean age 19.83; 7 female) took part in the 2-alternative forced-choice (2-AFC) training portion of the study. Data from this cohort of participants represents a subset of that obtained for a separate study (Powers et al., 2009). Because this study was designed to investigate the influence of simultaneity judgment training on perception of the SIFI, the criteria for inclusion in this cohort were based upon 1) the successful completion of training on the simultaneity judgment paradigm (with evidence of temporal window narrowing), and 2) completion of pre-training and post-training SIFI assessments. By self-report, all participants had normal hearing and vision, and none had any history of neurological or psychiatric disorders. All procedures for all subject groups were approved by the Vanderbilt University Institutional Review Board (IRB).

2-AFC Simultaneity Judgment Assessment

In this task (**Fig. A.1a and b**), participants judged whether the presentation of an auditory stimulus and a visual stimulus was 'simultaneous' or 'non-simultaneous' by pressing 1 or 2, respectively, on a response box (Psychology Software Tools Response Box Model 200A). Participants were seated in a quiet and dark room 48 cm from a computer monitor. E-Prime 2.0 (2.0.1.109) was used to control all experiments.

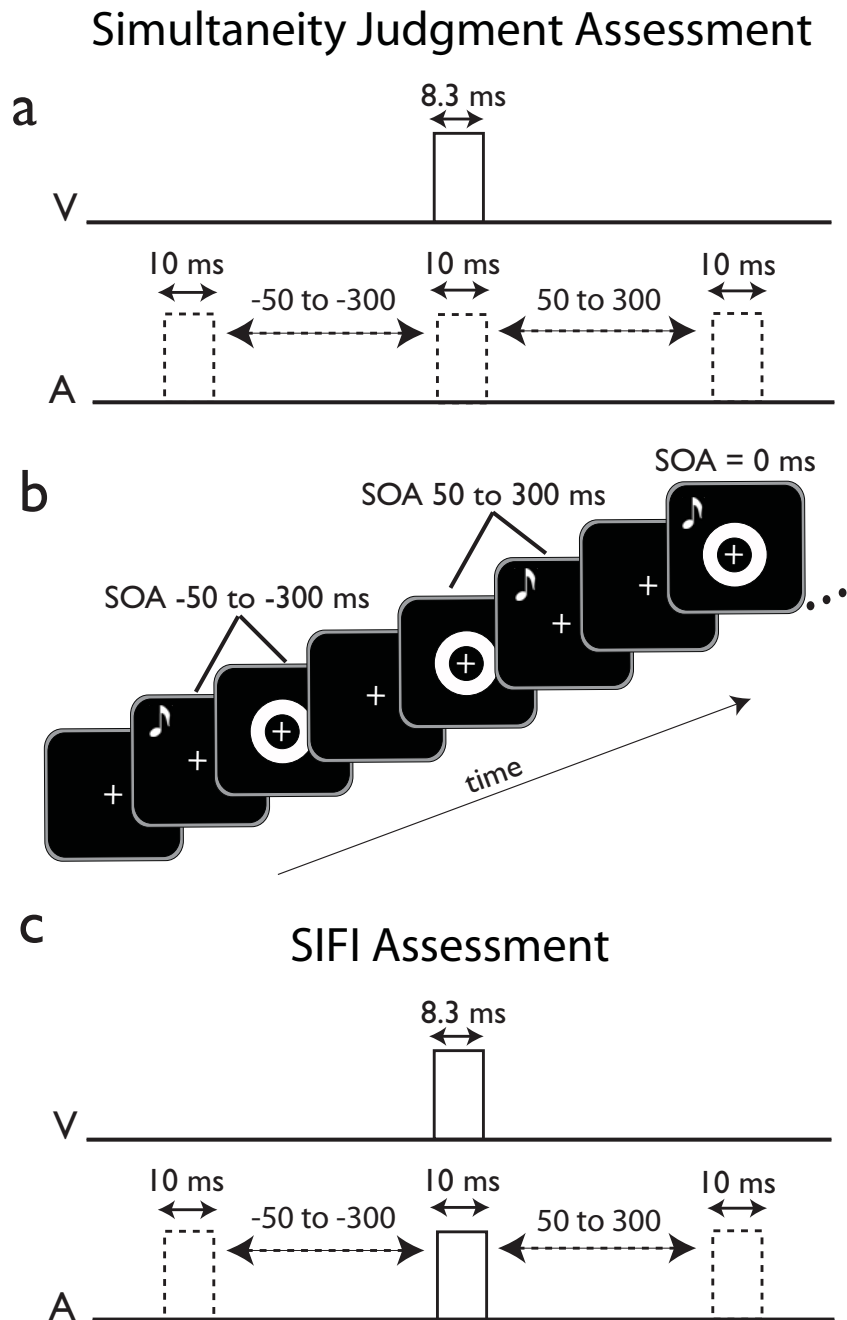


Figure A.1. Experimental Procedures. *a.* Temporal structure of the simultaneity judgment task. The 2-AFC version of the task consisted of one presentation of either a veridically simultaneous (SOA 0) or asynchronous (SOAs ranging from -300 ms to 300 ms by 50-ms increments) audiovisual stimulus pair, followed by a response period. The 2-IFC version presented both a simultaneous and an asynchronous pair per trial, followed by a response period. *b.* Schematic and characteristics of stimulus presentation. *c.* Temporal structure of the SIFI task, illusory (one-flash) condition. In this condition, one flash (a solid white circle eccentrically presented below the fixation cross) is accompanied by two tones, one of which always appears simultaneously with the flash. In the two-flash condition, flashes are separated by 52 ms and the central beep is presented at the midpoint between the flashes.

A white crosshair fixation marker (1 cm x 1 cm) on a black background appeared 1 second before the stimuli were presented and persisted throughout the duration of each trial. The visual stimulus consisted of a white ring on a black background subtending 15° of visual space with an outer diameter of 12.4 cm and an inner diameter of 6.0 cm (area = 369.8 cm²). This stimulus was presented for one refresh duration on a high refresh-rate monitor (NEC MultiSync FE992, 120 Hz; one refresh = 8.3 ms).

The auditory stimulus was a 10-ms, 1800-Hz tone burst presented binaurally via headphones (Phillips SBC HN110) with no interaural level or time differences. The stimulus was calibrated with a Larson-Davis sound level meter (Model # 814) and was presented at 110.4 dB SPL unweighted using impulse detection and flat weighting settings.

The stimuli had stimulus onset asynchronies (SOAs) ranging from -300 ms (auditory preceding visual) to 300 ms (visual preceding auditory) at 50 ms intervals. All stimulus timing was verified externally with an oscilloscope within an error tolerance of 10 ms arising from the inherent temporal imprecision of the auditory presentation hardware and drivers. In the Simultaneity Judgment Assessment task, temporal offsets were equally distributed. A total of 325 trials made up the task (25 cycles x 13 trials/cycle).

Sound-Induced Flash Illusion (SIFI) Task

Participants completed the SIFI task (Shams et al., 2000, 2002; Rosenthal et al., 2009) once directly after the baseline simultaneity judgment assessment and once again after the completion of the final simultaneity judgment assessment on Day 5. In this task,

participants are presented with one or two flashes paired with zero, one, or two beeps. Flashes were 8.5 ms in duration, and consisted of a white circle with an area of 12.6 cm² presented on a black background one centimeter below the fixation cross. Beeps consisted of a temporally-ramped 5000 Hz pure tone of 8 ms duration. In the two flash/one beep condition, the flashes were separated by 50 ms and the beep was presented at the midpoint between the two flashes. Similarly, in the two flash/two beep condition, the timing between flashes remained constant and one beep was always presented at the midpoint between the two flashes, with the other preceding or following it by an SOA ranging from 50 to 300 ms. In the illusion-inducing one flash/two beeps condition, one beep always occurred simultaneously with the flash onset, while one either preceded or followed that onset by SOAs ranging from 50 to 300 ms. This condition typically induces the perception of two flashes although only one appears, and the strength of this illusion varies with SOA (Shams et al., 2002). Each of these conditions occurred an equal number of times so as not to introduce a response bias. After each trial, participants responded by button-press to indicate the number of flashes they had perceived. In the SIFI assessment task there were 300 total trials (10 cycles x 30 trials/cycle) with an equal distribution of each condition.

2-AFC Simultaneity Judgment Training

The training task differed from simultaneity judgment assessments in that after making a response, the subject was presented with either the phrase “Correct!” paired with a happy face, or “Incorrect” paired with a sad face corresponding to the correctness of their choice. These faces (happy = yellow, sad = blue, area = 37.4 cm²) were presented

for 500 ms in the center of the screen. The white ring and fixation were of the same size and duration as in assessment trials. Only SOAs between -150 and 150 ms, broken into 50 ms intervals, were used during the training phase. Additionally, in this phase the SOAs were unequally distributed: the veridical simultaneous condition had a 6:1 ratio to any of the other 6 non-simultaneous conditions. In this way there was an equal likelihood of simultaneous/non-simultaneous conditions, minimizing concerns about introducing a response bias. The training phase consisted of 120 trials (20 cycles x 6 trials/cycle). See **Figs. A.1a** and **A.1b** for illustrations of the temporal relationship between stimuli.

2-AFC Training Protocol

Training occurred over 5 hours (1 hour per day) during which participants took part first in a pre-training simultaneity judgment assessment, next in one SIFI assessment, then in 3 shorter simultaneity judgment training runs, followed by a post-training simultaneity judgment assessment. An additional baseline assessment was performed at the start of the study for each subject, followed by the typical training day; this was designed to detect any practice effects that may have resulted from completion of the assessment itself.

Follow-Up Assessment

After one week without training, a subset of the training cohort described above (n=7, 2 female; mean age = 20.43) returned to the lab and underwent one simultaneity judgment assessment and one SIFI assessment without any training.

2-AFC Exposure

Subjects

Twelve (12) Vanderbilt undergraduate and graduate students (mean age 19.72; 4 female) underwent the 2-alternative forced-choice (2-AFC) exposure portion of the study. As with the 2-AFC training group data, data from this cohort of participants represents a subset of that obtained for a separate study (Powers et al., 2009). Inclusion in this cohort was based upon the successful completion of both the exposure protocol and the pre- and post-exposure SIFI assessments. All participants had self-reported normal sight and hearing, and none had any personal or family history of neurological or psychiatric disorders.

Exposure Protocol

The exposure portion of the study differed from the 2-AFC training protocol only in that instead of the training blocks, participants underwent 2-AFC exposure blocks of the same length. Thus, all participants in both cohorts took part in the same number of 2-AFC simultaneity judgment and SIFI assessments. The details of the exposure sessions are outlined below.

2-AFC Exposure

In the interest of maintaining attention, the 2-AFC exposure blocks were designed as an oddball task wherein participants were exposed to the same audiovisual pairs used in the simultaneity judgment training sessions but were instructed to press a button when they saw a red ring. As in the simultaneity judgment training sessions, the veridical

simultaneous condition had a 6:1 ratio to any of the other 6 non-simultaneous conditions. Oddballs occurred with the same probability across all conditions, and were 1/10 as likely to appear as the standard. The rings and fixation were of the same dimensions and duration as in the assessment trial; the tone was identical to that presented during the simultaneity judgment assessment and training sessions. A range of SOAs between -150 and 150 ms, in steps of 50-ms intervals, were used for this task.

2-IFC Training

Subjects

Thirteen (13) Vanderbilt undergraduate and graduate students (mean age 20.46; 10 female) underwent the 2-interval forced-choice (2-IFC) training portion of the study. As with data from the other cohorts, data from this cohort of participants represents a subset of that obtained for a separate study (Powers et al., 2009). Because this study was designed to investigate the influence of simultaneity judgment training on perception of the SIFI, inclusion in this cohort was based upon 1) the successful completion of training on the simultaneity judgment paradigm (with evidence of temporal window narrowing), and 2) completion of pre-training and post-training SIFI assessments. All participants had normal hearing and vision by self-report, and none had any personal or close family history of neurological or psychiatric disorders.

2-IFC Simultaneity Judgment Assessment

The 2-IFC simultaneity judgment assessment employed precisely the same stimuli as those used in the 2-AFC task. In this task, however, participants were presented with two audiovisual pairs, one with an SOA of zero (simultaneously-presented) and one with a non-zero SOA (non-simultaneously presented). Presentations were separated by 1 second, during which a fixation cross alone was presented. Participants were asked to indicate as quickly as possible by button-press which interval (first or second presentation) contained the flash and beep that happened at the same time. Simultaneous pairings were as likely to be presented in the first interval as in the second, and a simultaneous-simultaneous catch trial condition was present in equal representation to other SOAs.

2-IFC Simultaneity Judgment Training

The training phase of the 2-IFC portion of the study was identical to that of the assessment phase with two exceptions: 1) in the same manner described in the 2-AFC training, participants were given feedback as to the accuracy of their responses after each trial; 2) as in the 2-AFC simultaneity judgment training protocol, the range of SOAs presented during training (-150 ms to 150 ms by 50-ms increments) was restricted in training as compared to assessment (-300 ms to 300 ms). However, unlike the 2-AFC version of this training, the ratio of simultaneous to non-simultaneous presentation was always 1:1.

2-IFC Training Protocol

Participants underwent training in five 1-hour blocks (one hour per day) on the 2-IFC version of the simultaneity judgment task. Each day's 2-IFC training began with a simultaneity judgment assessment followed by three shorter blocks of training, and ended with a post-training simultaneity judgment assessment.

Follow-Up Assessment

A subset of the 2-IFC training cohort described above ($n=7$, 5 female; mean age = 20.57) returned to the lab one week after cessation of training and underwent one simultaneity judgment assessment and one SIFI assessment without any training.

Data Analysis

All data were imported from E-Prime 2.0 text files into MatLab 7.7.0.471 R2008b (The Mathworks, Inc., Natick, MA) via a custom-made script for this purpose. Individual subject raw data were used to calculate the mean probability of simultaneity judgment (2-AFC), accuracy (2-IFC), and proportion of trials at which two flashes were reported (SIFI) at each SOA for all assessments. These means were then analyzed in multiple ways as summarized in the following sections.

Estimation of Window Size

Mean data from each individual were fit with two sigmoid curves generated using the MatLab *glmfit* function, splitting the data into left (auditory presented first) and right (visual presented first) sides and fitting them separately. For the 2-AFC tasks, the

criterion at which to measure the breadth of the temporal window was equal to 75% of the maximum data point at baseline assessment. For the 2-IFC task, this criterion was set at half the distance between individuals' lowest accuracy point at baseline assessment and 1 (also ~ 75% accuracy). These criteria were then used to assess the breadth of the distributions produced by each individual's assessment data throughout the duration of the training period. Distribution breadth was then assessed for both the left side (from zero to the left-most point at which the sigmoid curve crossed the criterion line) and the right side (from zero to right intersection point) and then combined to get an estimation of total distribution width. This measure was then used as a proxy for the size of each individual's window at each assessment. An example of the result of this process may be seen in **Figure A.2a**. It should be noted that, when mean data from any individual assessment were unable to be fit with a sigmoid curve, all data from this individual were discarded for analysis of window size progression and SIFI performance change. Analysis of differences in window size across time was conducted by performing a repeated-measures ANOVA (within-subject factor, assessment number) followed by post-hoc t-tests (corrected via the Holm method for multiple comparisons) to determine which differences between assessment measures were responsible for the variance observed.

Signal Detection Analysis

In order to determine whether any changes in SIFI performance were the result of a true increase in perceptual sensitivity (d') or a shift in response bias (β), a signal detection analysis was performed. Perceptual sensitivity (d') was defined as the ability to

discriminate between one flash and multiple flashes (Green and Swets, 1966; Rosenthal et al., 2009). These parameters were calculated per individual in the following manner:

$$d' = z(H) - z(F)$$
$$\beta = 0.5 * (z(H) + z(F))$$

where $z(p)$ indicates the inverse of the cumulative normal distribution corresponding to the response proportion p . H (hit) denotes correct detection of multiple flashes, and F (false alarm) indicates an incorrect report of multiple flashes. In order to determine if either of these factors changed over the course of training or exposure, these values were calculated per individual on the basis of their pre- and post-training/exposure SIFI assessments and the mean group difference scores reported.

Results

Perceptual training on a simultaneity judgment task narrows the temporal window of multisensory binding

During the course of the multisensory perceptual learning paradigm described in (Powers et al., 2009), several subjects exhibited a narrowing of their multisensory temporal binding windows, as seen in an audiovisual simultaneity judgment task. Data produced by the two-alternative forced-choice (2-AFC) version of the training protocol from one participant are depicted in **Figure A.2a**. In this representation, the average probability of simultaneity judgment is plotted as a function of SOA. These data are then fitted with two sigmoid curves to model the left and right sides of the binding window. The breadth of this window was taken to be the length in time (in ms) at which individuals judged simultaneity at or greater than 75% of their highest baseline value.

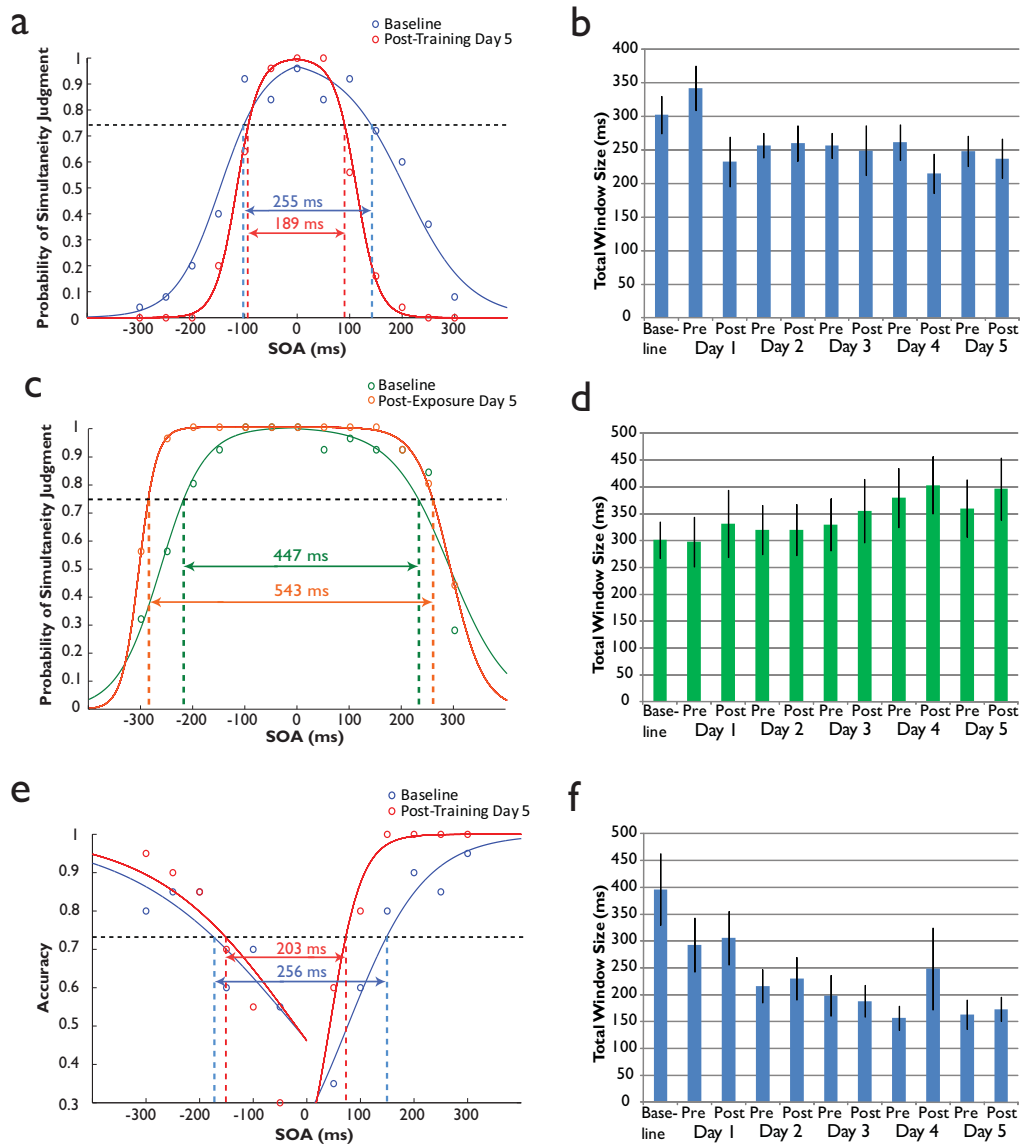


Figure A.2. Perceptual training on a simultaneity judgment task narrows the temporal window of multisensory binding. *a.* Assessment of window size and window size narrowing in one 2-AFC training participant. Average individual probability of simultaneity judgment is plotted as a function of stimulus onset asynchrony (SOA). Sigmoid curves are fit to either side of the resulting distribution, and the window breadth is taken as the width of this distribution at three quarters of the maximum data point at baseline. This individual’s window size at baseline is 255 ms by this measure, and narrows to 189 ms. *b.* Mean window size change over the course of 5 training days for 12 2-AFC training participants. Error bars represent one SEM. *c.* Individual window size derivation from one exposure participant at baseline and post-exposure assessment on day 5. *d.* Progression of window size in 13 exposure control participants. Note that these participants do not exhibit the decrease characteristic of training group participants, but actually show a trend toward window size expansion. *e.* Individual derivation of window size in one 2-IFC training participant. Note that in the 2-IFC version of the task data is plotted as accuracy versus SOA. In a similar manner to that used in the 2-AFC version of the task, these data are fit with two sigmoid curves and window size is taken as the halfway point between the minimum data point at baseline and perfect accuracy (~75%). *f.* Mean total window size change over 5 days of training on the 2-IFC task. Note that, as with the 2-AFC subjects, window size narrows by nearly half compared to baseline assessment.

The mean window size for the individual data depicted here is shown to be 255 milliseconds.

After training on the simultaneity judgment paradigm there was a dramatic shift in the size of the multisensory temporal binding window. In the individual shown, this shift translated to a 26% reduction in the size of the window, to 189 ms. To examine the time course of this narrowing on the group level, we plotted each individual's distributions at each of the 11 assessments, derived window sizes from these plots, and then charted the mean window size over the course of the 5 days of training. This time course is depicted in **Figure A.2b** for 12 subjects whose window sizes decreased from baseline to the post-training Day 5 assessment (described as dynamic subjects in Powers et al., 2009). A repeated-measures ANOVA revealed significant narrowing in these subjects' temporal windows over the course of training ($F_{10,90} = 3.6798$; $p = 0.00038$) with mean window sizes decreasing significantly from 302 ms at baseline assessment to 237 ms at the post-training assessment on Day 5 (by paired-samples t-tests, $p = 0.0391$, corrected). It should be noted that this outcome is not unexpected, as this cohort was specifically chosen to examine the effect of window narrowing on performance of an unrelated multisensory temporal task.

The results above may be contrasted with those of the exposure group, depicted in **Figure A.2c** and **A.2d**. As with the training group, individual assessment data from the exposure group ($n = 9$) were fitted with two sigmoid curves and window sizes were measured by the same 75% maximum criterion (**Fig. A.2c**). Surprisingly, results indicated an increase in window size over the course of the week of exposure (by repeated-measures ANOVA, $F_{10,80} = 2.212$; $p = 0.0250$), with window size first reaching

significance over baseline assessment (301 ms) at pre-exposure assessment on Day 4 (380 ms; by paired samples t-test, $p = 0.0266$, corrected) and remaining significantly larger upon final assessment (396 ms; $p = 0.0153$, corrected).

To control for the possibility that the changes in window size described in the training group above may have been driven by changes in cognitive biases (i.e., criterion shifts), rather than by a true change in perceptual processes, several subjects were run on a two-interval forced-choice (2-IFC) version of the task, wherein subjects are requested to judge which of two sequential presentations contained the simultaneous audiovisual pair. This structure does not require the setting of a criterion for simultaneity and thus is more likely to reveal true differences in discrimination ability. Individual assessment data in the form of mean accuracy per SOA were plotted and sigmoid curves fitted to each side of the distribution, just as in the 2-AFC version of the task (**Fig. A.2e**). The size of the temporal window here was defined as halfway between each individual's lowest accuracy point at baseline and 1 (the mean criterion level was 72%), and this measurement was derived for each of the 11 assessments over 5 days of training. Results once again showed a narrowing of group window size (see Powers et al., 2009 for full cohort data). **Figure A.2f** summarizes data from those subjects whose window sizes were narrowed with training (repeated-measures ANOVA, $F_{10,110} = 4.2568$; $p = 5.13 \times 10^{-3}$) and who were assessed on the SIFI task as well ($n=13$). As expected, total window size in these subjects narrowed significantly from 396 ms at baseline to 173 ms at post-training Day 5 assessment ($p = 1.5 \times 10^{-4}$, corrected). This narrowing appeared very early during the training week, at the pre-training assessment on Day 2 ($p = 2.0 \times 10^{-4}$, corrected).

Both 2-AFC and 2-IFC training result in sensitivity increases upon signal detection analysis of SIFI results

Participants in both the training and passive exposure groups took part in the SIFI task immediately after the baseline simultaneity judgment assessment (i.e., prior to training) and again after the final simultaneity judgment assessment following the 5 days of training. An analysis of the data using signal detection theory (SDT) was undertaken to determine whether performance on the SIFI task changed in each of the three groups. **Figure A.3a** highlights the difference in sensitivity (d') after one week of training or exposure in each group. While both 2-AFC (mean = 0.564; $p = 0.038$ by paired t-test) and 2-IFC (mean = 0.604; $p = 0.0036$) training groups exhibited marked increases in mean sensitivity between pre- and post-training assessments, the exposure group showed no such difference (mean = -0.0680; $p = 0.78$). By contrast, analysis of the response bias term (β ; **Fig. A.3b**) revealed very small pre/post differences in means for all three groups (2-AFC training: 0.0929; 2-AFC Exposure: 0.0712; 2-IFC training: 0.213), although this difference for the 2-IFC group did reach statistical significance ($p = 0.0428$ by paired-samples t-test).

To determine whether the changes observed in the d' measure were driven by increases in hits or decreases in false alarms in these individuals (see equation for d' calculation in Materials and Methods), changes in the proportions of hits and false alarms in each of the three groups were analyzed and plotted (**Fig. A.3c**). Results indicate that, while there was a decrease in the mean proportion of false alarms for the 2-AFC training group (mean = -0.0258), this difference was not statistically significant upon paired-samples t-test ($p = 0.315$), and no other groups exhibited appreciable differences on this

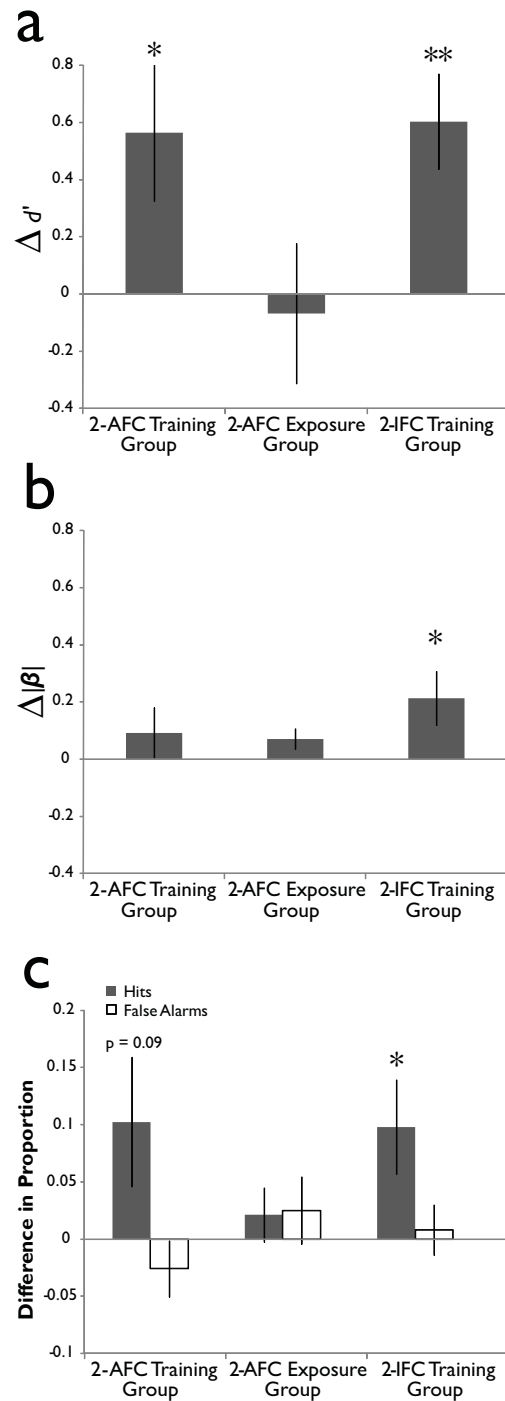


Figure A.3. Both 2-AFC and 2-IFC training result in sensitivity increases upon signal detection analysis of SIFI results. *a.* Change in sensitivity (d') after training/exposure in all three groups. In striking contrast to the exposure group, both training groups exhibit a significant increase in sensitivity after training on the simultaneity judgment task. *b.* Change in response bias (β) after training/exposure. Note the difference in magnitude of change when compared to sensitivity shifts. *c.* Breakdown of factors contributing to the sensitivity shift shown in *a.* While the 2AFC training group does exhibit a small decrease in proportion of false alarms after training, the sensitivity differences seen appear to be driven primarily by increases in hits, or correct identifications of two-flash presentations. Error bars indicate one SEM; * $p < 0.05$; ** $p < 0.01$.

measure. In stark contrast, both the 2-AFC and 2-IFC training groups showed marked increases in hit proportions from pre- to post-training assessments (~10% for each; $p = 0.098$ and $p = 0.0346$, respectively). Thus, the changes in performance after training were the result of increases in correct recognition of two closely-presented flashes rather than a decrease in the proportion of trials over which the SIFI was reported.

Changes induced by perceptual training on both the 2-AFC and 2-IFC simultaneity judgment tasks are stable for at least 1 week

One week after completion of training, 7 subjects from the 2-AFC training group described above returned to the lab for follow-up simultaneity judgment and SIFI assessments. **Figure A.4a** depicts the progression of window size narrowing for this follow-up group. Note that window sizes for these subjects at 1-week follow-up assessment (mean of 227 ms) remain significantly smaller than at baseline assessment (mean = 314 ms; $p = 0.045$ by paired-samples t-test), but do not differ significantly from the window size as measured at post-training Day 5 assessment (221 ms, $p = 0.957$).

Similarly, a group of 7 subjects from the 2-IFC training group above returned for follow-up assessments (**Fig. A.4b**). Again, results reveal a window size that remains narrow upon 1-week follow-up (164 ms) when compared with baseline assessment (414 ms, $p = 0.0095$), but does not differ significantly from the window size at post-training Day 5 assessment (192 ms, $p = 0.110$).

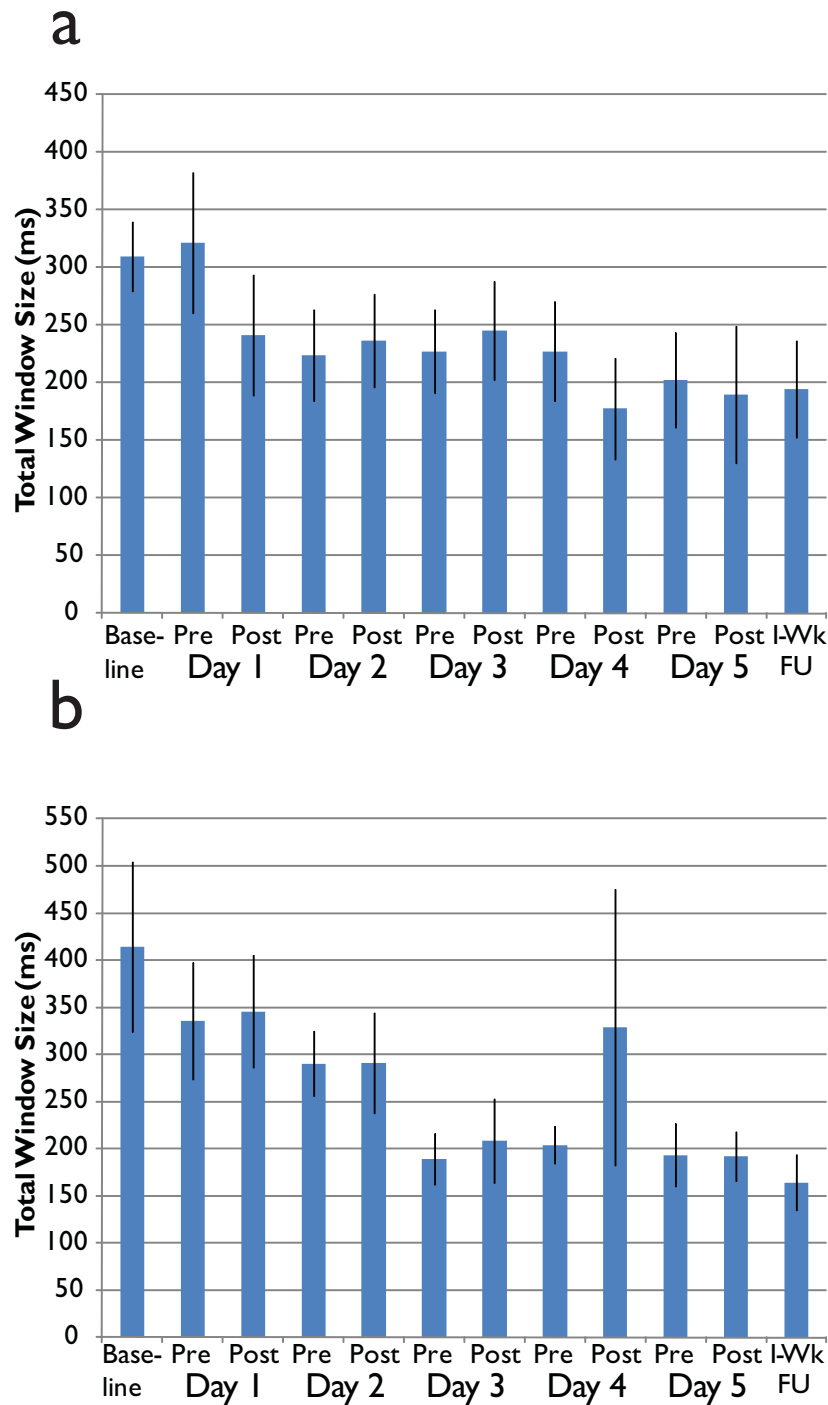


Figure A.4. Changes induced by perceptual training on both the 2-AFC and 2-IFC simultaneity judgment tasks are stable for at least 1 week. **a.** Average progression of window sizes in a cohort of 7 participants from the 2-AFC training group shown in **Figure 2a** upon assessment one week after cessation of training. **b.** Progression of mean window sizes in a cohort of 7 participants from the 2-IFC group upon 1-week follow-up. Error bars indicate one SEM.

Increases in sensitivity on the SIFI task after training are stable for at least 1 week

Because the effects of training show evidence of persistence one week after training cessation, the longevity of the changes in SIFI performance was also assessed. As summarized in **Figure A.5a**, both training groups continued to exhibit large mean increases in sensitivity from baseline (2AFC: 0.990; 2-IFC: 0.643), although these changes only approached statistical significance ($p = 0.0548$ and $p = 0.0556$, respectively). Once again, there appeared to be small but consistent increases in bias upon follow-up assessment (**Fig. A.5b**; 2-AFC mean: 0.118; 2-IFC mean: 0.211), and the 2-IFC increase reached statistical significance ($p = 0.0247$).

As was done with the data acquired immediately after training, the changes in sensitivity on the SIFI task were broken down in terms of changes in the proportions of hits and false alarms between pre-training and 1-week follow-up assessments (**Fig. A.5c**). Once again, the effects seemed to be driven primarily by increases in the proportion of hits over this interval (mean change 2-AFC = 0.186; mean change 2-IFC = 0.084), although neither of these changes reached significance ($p = 0.113$ and $p = 0.144$, respectively). This is most likely due to the decreased number of participants returning for follow-up. In contrast, there was no appreciable change in proportion of false alarms in either training group (2-AFC mean = -0.055; 2-IFC mean = -0.0120).

Increases in sensitivity correlate with the degree of window narrowing in the 2-AFC training group

To determine whether the degree of window size change brought about by training is able to predict the degree of change in sensitivity on the SIFI task, individual

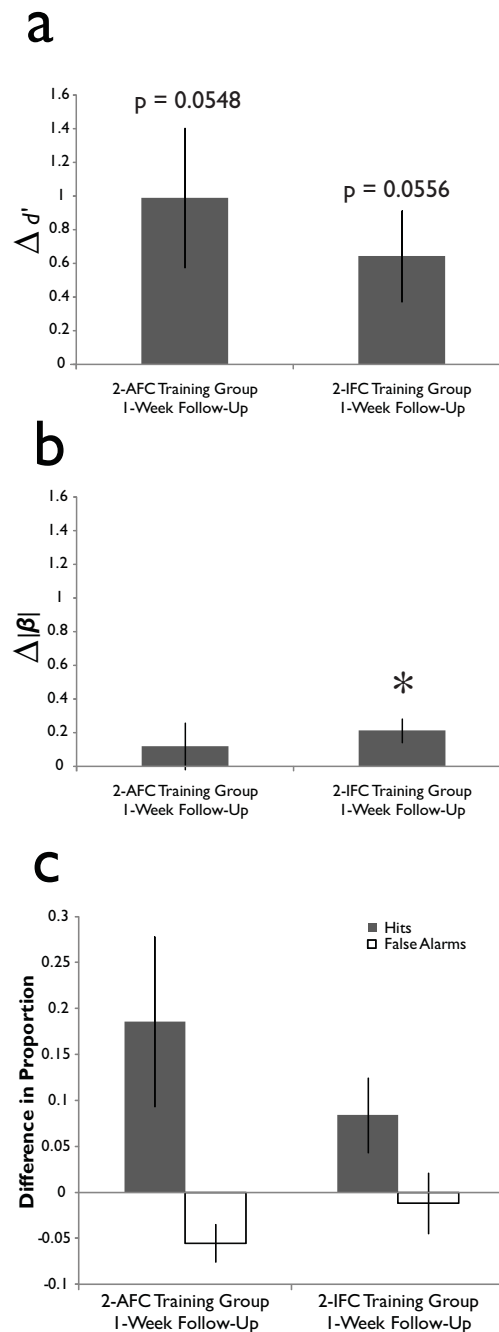


Figure A.5. Increases in sensitivity on the SIFI task after training are stable for at least 1 week. *a.* Change in sensitivity (d') from baseline SIFI assessment from baseline to one-week follow-up in both training groups. Both groups exhibit large increases in mean sensitivity upon one-week follow-up assessment, although both only approach significance. *b.* Change in response bias (β) in these participants. Once again, note the difference in magnitude between this change and the sensitivity shift shown in *a.* *c.* Breakdown of factors influencing the sensitivity shift described. As seen in the assessment immediately following training, this shift appears to be primarily driven by an increase in hits, although smaller decreases in false alarms are also present. Error bars indicate one SEM; * $p < 0.05$.

difference scores in window size and d' were entered as factors into a linear regression. **Figure A.6** illustrates the result of this procedure for the 2-AFC training group. As seen in **Figure A.6a**, there is a direct correlation between the percent decrease in window size exhibited by members of this group and the percent increase in sensitivity seen on the SIFI task in these subjects immediately after training ($r^2 = 0.4133$; $p = 0.0145$). Analysis of the 2-AFC follow-up data (**Fig. A.6b**) revealed a similar relationship, indicating a positive correlation between degree of change seen on the simultaneity judgment task and the SIFI task ($r^2 = 0.793$; $p = 0.0078$). In contrast, no significant correlations between these measures were noted in the 2-AFC Exposure or in the 2-IFC training group.

Discussion

We have shown that individuals who have undergone perceptual training on an audiovisual simultaneity judgment task exhibit altered performance on a sound-induced flash illusion (SIFI) task. Further, we have shown that this change in performance is primarily driven by an increase in recognition of two-flash conditions after training. Moreover, we have demonstrated that the magnitude of change in performance on the SIFI task is directly dependent upon the degree of temporal window narrowing wrought by the simultaneity judgment training paradigm. Finally, we have shown that these changes are stable over time and remain robust at least one week following the cessation of training.

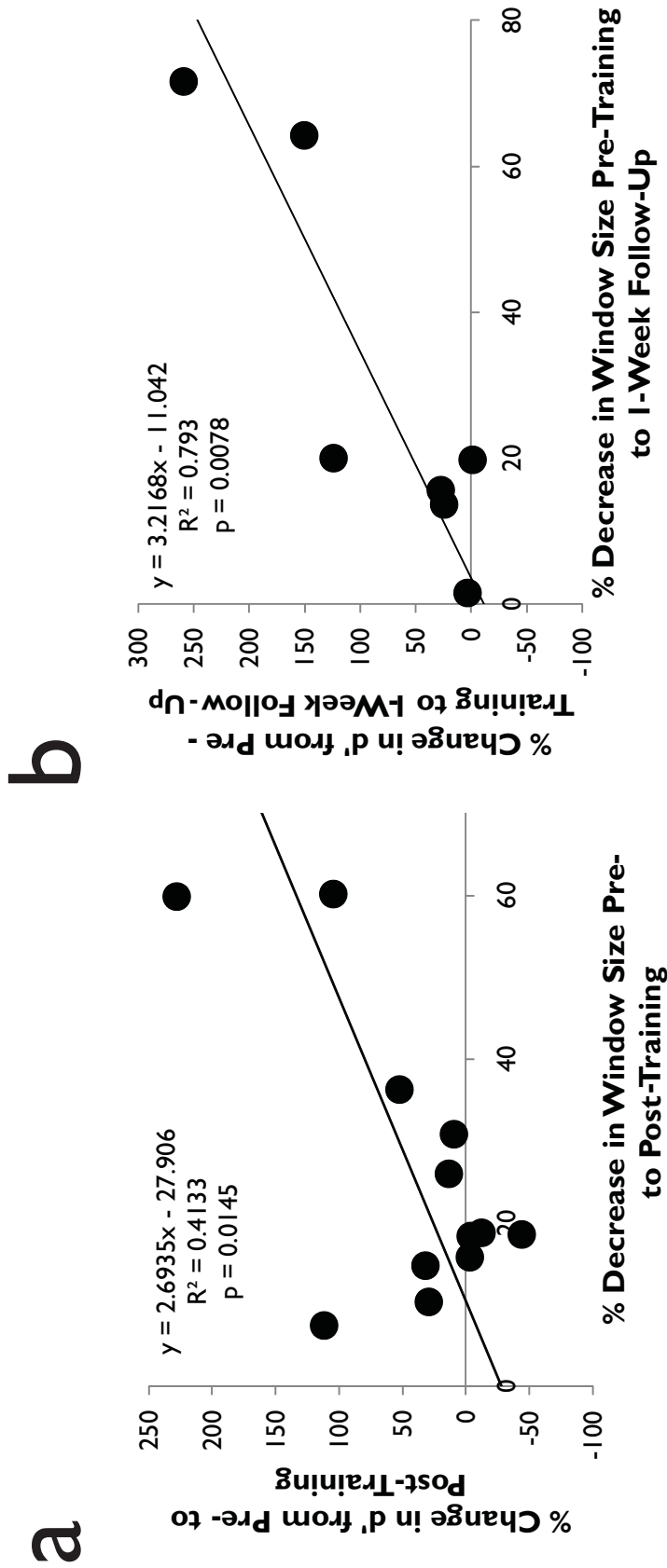


Figure A.6. Increases in sensitivity correlate with the degree of window narrowing in the 2-AFC training group. a. Relationship between magnitude of window narrowing exhibited in the simultaneity judgment task and percent change in sensitivity from pre- to post-training in all 2-AFC training participants. b. Correlation between window size change from baseline to 1-week follow-up assessment and percent change in sensitivity in the 7 2-AFC participants.

The degree to which the changes observed on the SIFI task are related to the alterations in window size observed in the primary simultaneity judgment task is important to the interpretation of the results outlined above. **Figure A.6** establishes a relationship between the degrees of change exhibited by 2-AFC training subjects in the two tasks, but it should be noted that no such relationship could be discerned for the 2-IFC training group. Possible reasons for this seeming incongruence include the apparent dependence of the 2-AFC effects not only on an increase in hits after training, but on a decrease in false alarms. **Figures A.7** and **A.8** show a correlation between the degree of temporal window narrowing and the decrease in false alarms shown by participants in all three groups in post-training and 1-week follow-up assessments. Most striking in this analysis is that a strong relationship is seen in the 2-AFC training group but remains completely absent in both the 2-AFC exposure group and the 2-IFC training group, supporting the idea that the correlation between the sensitivity measure and window size narrowing shown in **Figure A.6** is likely driven not by the robustly demonstrated increase in hits, but by the decrease in false alarms exclusively seen in the 2-AFC training group. The fact that this difference exists between the two training groups is perplexing, but because a single-presentation 2-AFC task of this type (as opposed to the 2-IFC construct) relies upon the setting of an internal criterion (Nachmias, 1981; Pelli, 1985), a post-perceptual mechanism for this decrease is likely. Indeed, the fact that the two groups are so similar in terms of window narrowing on the simultaneity judgment task and in terms of their increase in hits on the SIFI task points to both perceptual and post-perceptual changes in the 2-AFC group on both tasks. The presence of such top-down influences on the illusion itself (here described as a decrease in false alarms) is unprecedented in the

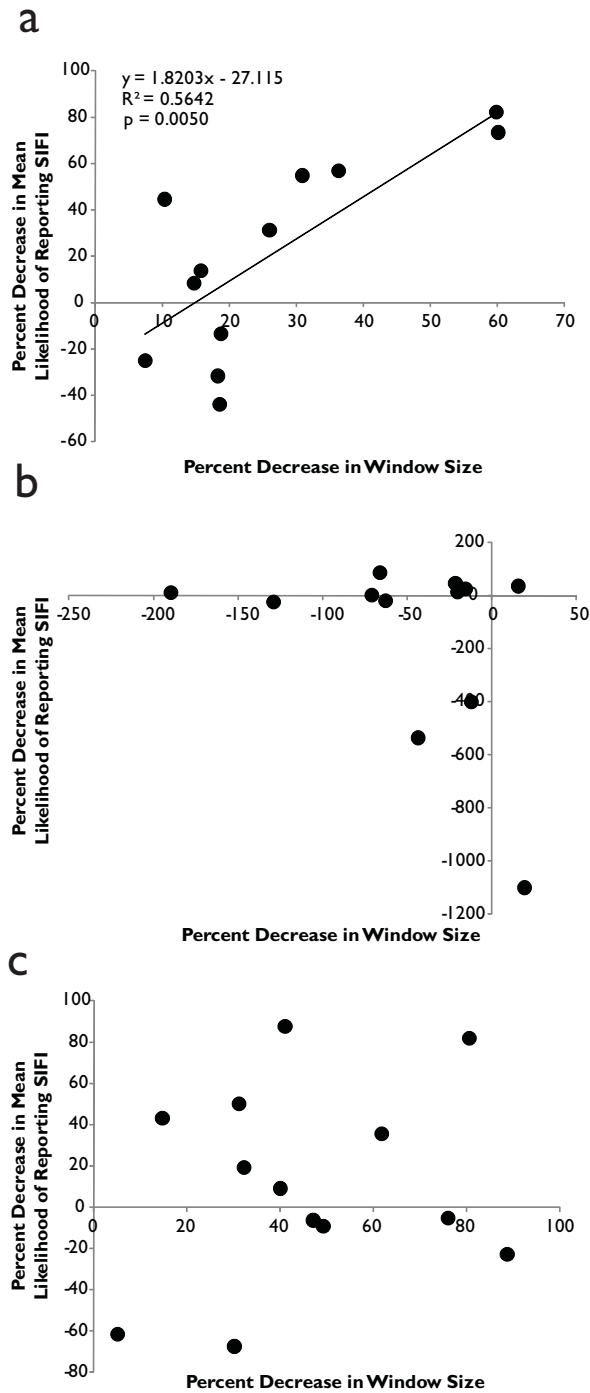
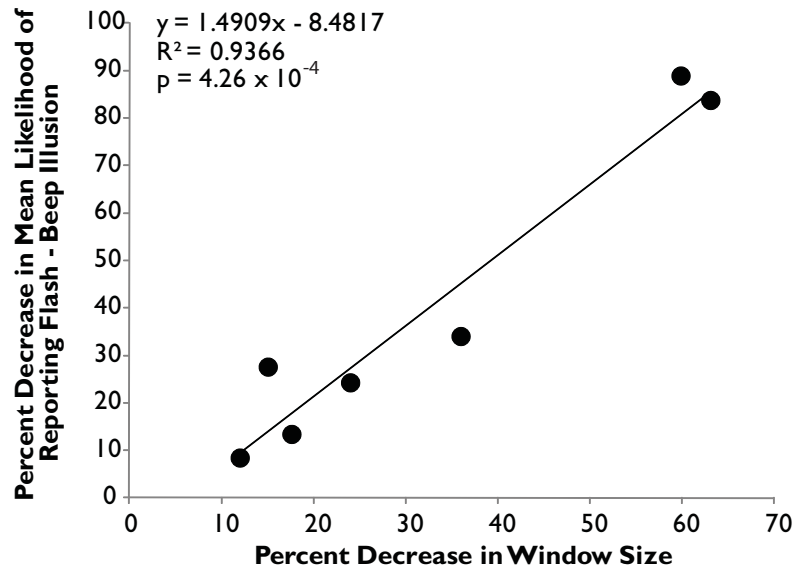


Figure A.7. Degree of window size narrowing and percent decrease in propensity to report the SIFI covary only in the 2-AFC training group. Relationship between percent window size decrease and percent decrease in likelihood of reporting the SIFI in 2-AFC participants (a), 2-AFC exposure participants (b), and 2-IFC training participants (c).

a



b

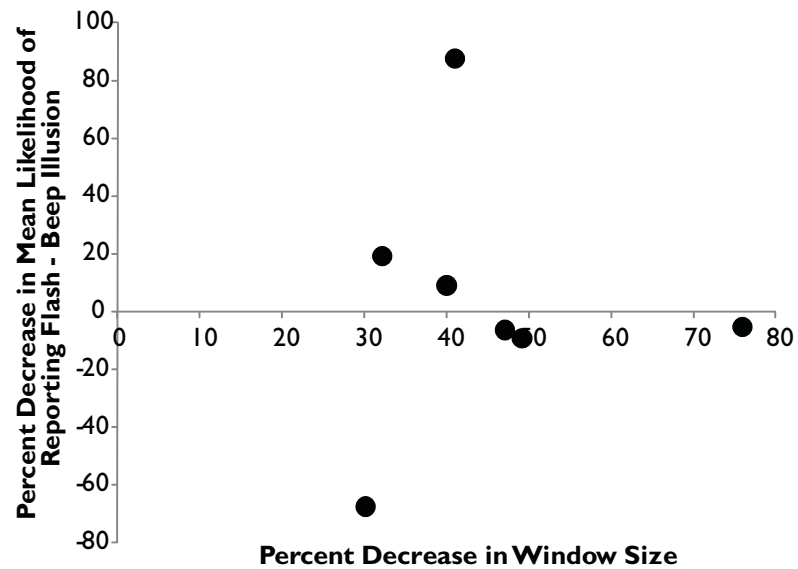


Figure A.8. Degree of window size narrowing and percent decrease in propensity to report the SIFI covary only in the 2-AFC training group upon one-week follow-up assessment. Relationship between percent window size decrease and percent decrease in likelihood of reporting the SIFI in 2-AFC participants (a) and 2-IFC training participants (b) on one-week follow-up assessment.

literature. In fact, the illusion's seeming inability to be altered by feedback training (Rosenthal et al., 2009) is suggestive of its imperviousness to top-down influences. Nonetheless, the data reported here do demonstrate a difference in the probability of individuals to report the SIFI that clearly depends upon the amount of change induced by the primary task after training, arguing strongly that participants' propensities to report the SIFI are indeed susceptible to post-perceptual influences.

The interrelatedness of the simultaneity judgment and SIFI tasks in terms of their temporal structures (see **Fig. A.1**) prompts the question as to whether, as in the original training task, differences in SIFI performance may be limited to a specific range of stimulus onset asynchronies (SOAs). As demonstrated in **Figure A.2** and as highlighted in Powers (2009), the window narrowing observed after training is seen primarily as performance increases in middle to large lags on the right side of the distribution, corresponding to conditions wherein visual events precede their auditory counterparts. **Figure A.9** breaks down the pre-/post-training sensitivity changes by SOA in all three groups. In both the 2-AFC and 2-IFC training group—but not in the 2-AFC exposure group—sensitivity increases are in large part isolated to positive SOAs and at the intervals that show the most change in the simultaneity judgment task. While the nature of the two tasks is not similar enough to warrant a strict SOA-by-SOA prediction of generalization of effects between the two tasks, the fact that the same lateralization of effect exists in both supports their interrelatedness.

The effects here are driven primarily by increases in participants' abilities to discriminate between the presentation of one flash versus two flashes presented in close temporal proximity. Thus, the effects seen are essentially increases in performance

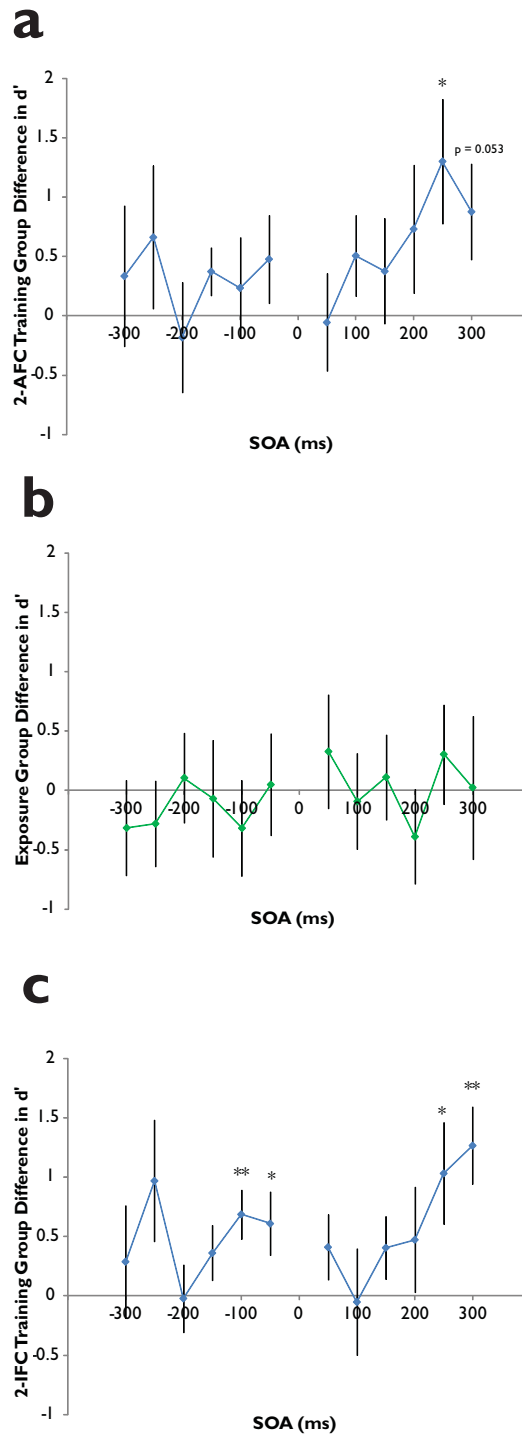


Figure A.9. Changes in sensitivity on the SIFI task after training occur mostly in positive SOA conditions. **a.** Mean differences in sensitivity as a function of SOA on the SIFI task for 2-AFC training participants. Note the marked improvement on the right side of the distribution, resembling the pattern of improvement on the simultaneity judgment task (Fig. A.2). **b.** Mean differences in sensitivity as a function of SOA in 2-AFC exposure participants. **c.** Mean differences in sensitivity by SOA on the SIFI task in 2-IFC training participants. Note the similarity of this pattern with that of the changes seen in the 2-AFC training group, and their seeming isolation to positive SOAs.

accuracy on a visual gap detection task. Other investigations into cross-modal generalization of temporally-based perceptual learning have generated mixed results: transfer of learning has been shown from training on a somatosensory timing task to a corresponding auditory task if similar intervals are tested (Nagarajan et al., 1998), but others have failed to demonstrate transfer of simultaneity learning within modalities and cross-modally (Virsu et al., 2008). While the results reported here do not necessarily support the existence of a single, crossmodal clock in its classical formulation as a pacemaker-accumulator (Treisman, 1963), they do join others (Alais and Burr, 2003; Burr and Morrone, 2006; Burr et al., 2009) in demonstrating the possible existence of shared components for timing perception among the sensory modalities. Indeed, these results fit well with a growing literature in support of interval-specific timing circuits that are dependent upon the time scale in question but independent of stimulus specifics, location, or modality (Ivry and Spencer, 2004; Buhusi and Meck, 2005; Johnston et al., 2006). Along these lines, future investigations should focus upon whether, as these results and cue reliability models of multisensory integration may predict (Deneve and Pouget, 2004; Andersen et al., 2005; Roach et al., 2006; Angelaki et al., 2009; Ronsse et al., 2009), the relationship between the multisensory and visual improvements described here may be causally linked.

References

- Alais D, Burr D (2003) The "Flash-Lag" effect occurs in audition and cross-modally. *Curr Biol* 13:59-63.
- Andersen TS, Tiippana K, Sams M (2005) Maximum Likelihood Integration of rapid flashes and beeps. *Neurosci Lett* 380:155-160.
- Angelaki DE, Gu Y, DeAngelis GC (2009) Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol* 19:452-458.
- Buhusi CV, Meck WH (2005) What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* 6:755-765.
- Burr D, Morrone C (2006) Time perception: space-time in the brain. *Curr Biol* 16:R171-173.
- Burr D, Silva O, Cicchini GM, Banks MS, Morrone MC (2009) Temporal mechanisms of multimodal binding. *Proc Biol Sci* 276:1761-1769.
- Colonus H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci* 16:1000-1009.
- Colonus H, Diederich A, Steenken R (2009) Time-window-of-integration (TWIN) model for saccadic reaction time: effect of auditory masker level on visual-auditory spatial interaction in elevation. *Brain Topogr* 21:177-184.
- Deneve S, Pouget A (2004) Bayesian multisensory integration and cross-modal spatial links. *J Physiol Paris* 98:249-258.
- Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys* 66:1388-1404.
- Diederich A, Colonius H (2008) Crossmodal interaction in saccadic reaction time: separating multisensory from warning effects in the time window of integration model. *Exp Brain Res* 186:1-22.

- Diederich A, Colonius H (2009) Crossmodal interaction in speeded responses: time window of integration model. *Prog Brain Res* 174:119-135.
- Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9:719-721.
- Doshier BA, Lu ZL (2007) The functional form of performance improvements in perceptual learning: learning rates and transfer. *Psychol Sci* 18:531-539.
- Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. New York,: Wiley.
- Ivry RB, Spencer RM (2004) The neural representation of time. *Curr Opin Neurobiol* 14:225-232.
- Jeter PE, Doshier BA, Petrov A, Lu ZL (2009) Task precision at transfer determines specificity of perceptual learning. *J Vis* 9:1 1-13.
- Johnston A, Arnold DH, Nishida S (2006) Spatially localized distortions of event time. *Curr Biol* 16:472-479.
- Lapid E, Ulrich R, Rammsayer T (2009) Perceptual learning in auditory temporal discrimination: no evidence for a cross-modal transfer to the visual modality. *Psychon Bull Rev* 16:382-389.
- Lewkowicz DJ (1996) Perception of auditory-visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform* 22:1094-1106.
- McGrath M, Summerfield Q (1985) Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J Acoust Soc Am* 77:678-685.
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215-3229.
- Nachmias J (1981) On the psychometric function for contrast detection. In, pp 215-223.

Nagarajan SS, Blake DT, Wright BA, Byl N, Merzenich MM (1998) Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J Neurosci* 18:1559-1570.

Pandey PC, Kunov H, Abel SM (1986) Disruptive effects of auditory signal delay on speech perception with lipreading. *J Aud Res* 26:27-41.

Pelli DG (1985) Uncertainty explains many aspects of visual contrast detection and discrimination. In, pp 1508-1531: OSA.

Polat U (2009) Making perceptual learning practical to improve visual functions. *Vision Res* 49:2566-2573.

Powers AR, 3rd, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. *J Neurosci* 29:12265-12274.

Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc Biol Sci* 273:2159-2168.

Ronsse R, Miall RC, Swinnen SP (2009) Multisensory integration in dynamical behaviors: maximum likelihood estimation across bimanual skill learning. *J Neurosci* 29:8419-8428.

Rosenthal O, Shimojo S, Shams L (2009) Sound-induced flash illusion is resistant to feedback training. *Brain Topogr* 21:185-192.

Roth DA, Appelbaum M, Milo C, Kishon-Rabin L (2008) Generalization to untrained conditions following training with identical stimuli. *J Basic Clin Physiol Pharmacol* 19:223-236.

Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. *Nature* 408:788.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14:147-152.

Treisman M (1963) Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychol Monogr* 77:1-31.

Virsu V, Oksanen-Hennah H, Vedenpaa A, Jaatinen P, Lahti-Nuuttila P (2008) Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Exp Brain Res* 186:525-537.