

The Plasticity of Temporal Perception: Perceptual Training Enhances Multisensory Temporal

Acuity

By

Matthew Allen De Niar

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 2016

Nashville, Tennessee

Approved:

James W. Bodfish, Ph.D., Committee Chair

Mark T. Wallace, Ph.D., Dissertation Advisor

Randolph R. Blake, Ph.D.

John C. Gore, Ph.D.

Tyler Reimschisel, M.D.

## **Dedication**

To my friends and family who are always with me

## Acknowledgements

I would like to thank all of the individuals, friends and family, who have supported me in the pursuit of my goals and ambitions and have guided my development as a person and scientist. You have all provided me with invaluable support, care, and love and I know that you will all continue to be an essential part of my life.

My growth as a scientist is attributable to numerous individuals. I would first like to thank the individuals who introduced me to neuroscience, particularly Dr. Richard Servatius, who allowed me to be part of his lab as a young high school student and has continued to encourage me throughout my career. I would also like to thank Dr. Julio Ramirez who not only supported me throughout my time at Davidson College but who has also continued to provide me with mentorship over the many years since leaving his lab for Vanderbilt. Finally, I would like to thank Dr. Mark Wallace for providing me with support and encouragement during the course of my graduate training. It has been an amazing opportunity to develop and complete the following projects as a part of the Wallace lab and I know that all I have learned during this time will continue to be valuable to me throughout my career.

I would also like to thank my friends and colleagues for their assistance in this project. In particular, I would like to thank Justin Siemann for helping me navigate the various phases of graduate school, Jean-Paul Noel for his constant collaboration on numerous projects and thoughtful scientific discussions, David Simon for helping develop numerous experiments, Aaron Nidiffer for teaching me how to use Matlab, Walter Lee for solving numerous technical and logistical problems, and Sarah Baum for her assistance in the analysis of the neuroimaging components of this project. I would also like thank all of the members of the Wallace lab who have been amazing colleagues and fellow students. Additionally, I would like to especially thank all of the undergraduate students who have worked with me over the past years as you have

provided exceptional assistance in the development and implementation of many of these projects.

I am also fortunate to have had the support of the Vanderbilt MSTP, the graduate school, and my dissertation committee during this process. I'd like to thank Dr. Terry Dermody for his guidance as director of the MSTP program and support during my graduate training, Melissa Krasnove for helping me navigate all the challenges and requirements as MSTP student, and Rosalind Johnson and Beth Sims for their help planning numerous academic events and logistical assistance. I'd also like to thank the members of my committee for providing valuable insight and knowledge as well as support and assistance. It has truly been a privilege to share my work with you and for you to share your expertise with me.

Finally, I would like to thank my family for their support over the course of my entire life and the opportunities they have provided me with. To my parents, Allen and Linda, thank you for supporting in all of my life's endeavors and your continued investment in my education and intellectual growth. To my sister, Marilena, thank you for being a wonderful sibling and providing me with numerous baked goods to continue studying. To Addy, thank you for being a faithful dog and sitting with me as I studied for many tests and wrote this dissertation. To Grace, thank you for all that you have given to me to support my career and myself as human being. I can truly say that without you this achievement would not be possible and I look forward to the many opportunities and blessings to come throughout our lives. You are a loving friend.

## Table of Contents

	Page
DEDICATION.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF FIGURES.....	xii
LIST OF TABLES.....	xiv
Chapter	
<b>I. Introduction.....</b>	<b>1</b>
An Introduction to Multisensory Processing.....	2
<i>Defining Multisensory Interactions.....</i>	<i>2</i>
<i>The Principles of Multisensory Processing.....</i>	<i>3</i>
<i>Influence of Multisensory Interactions on Behavior.....</i>	<i>5</i>
Temporal Aspects of Multisensory Processing.....	7
<i>Defining and Measuring Temporal Perception.....</i>	<i>7</i>
<i>The Relationship of Temporal Proximity, Synchrony Perception, and</i>	
<i>Multisensory Integration.....</i>	<i>14</i>
<i>Factors Influencing Temporal Perception.....</i>	<i>18</i>
<i>Neural Correlates of Multisensory Temporal Perception.....</i>	<i>24</i>
Multisensory Plasticity.....	26
<i>Development, Experience, and Plasticity in Multisensory Model Neural</i>	
<i>Systems.....</i>	<i>27</i>
<i>Multisensory Recalibration.....</i>	<i>32</i>
<i>An Introduction to Perceptual Learning.....</i>	<i>36</i>
<i>Concepts and Models of Perceptual Learning.....</i>	<i>37</i>
<i>Feedback and Perceptual Learning.....</i>	<i>40</i>

<i>Generalization of Perceptual Learning</i> .....	43
<i>Multisensory Perceptual Learning and Temporal Acuity</i> .....	47
<i>Neural Evidence of Perceptual Learning</i> .....	50
Introduction to the Subsequent Chapters .....	52
References.....	55
<b>II. The Effects of Feedback Type and Reliability on Multisensory Perceptual Learning .</b>	<b>74</b>
Abstract.....	74
Introduction .....	74
Material and methods .....	78
<i>Subjects</i> .....	78
<i>Simultaneity Judgment (SJ) Assessment</i> .....	78
Experiment 1 .....	79
<i>Protocol</i> .....	79
<i>Simultaneity Judgment (SJ) Training Session</i> .....	80
Experiment 2.....	81
<i>Protocol</i> .....	81
Simultaneity Judgment (SJ) Training Session .....	81
<i>Exposure Session</i> .....	82
Analysis.....	82
<i>Grand mean SOA analysis</i> .....	82
<i>Estimation of the TBW</i> .....	83
Results: Experiment 1 .....	83
<i>Decreasing the reliability of feedback impairs multisensory perceptual learning</i> .....	83
Results: Experiment 2.....	85
<i>Perceptual Training with Symbolic, Social, and Monetary Feedback on a</i>	

<i>SJ Task Enhances Multisensory Temporal Acuity</i> .....	85
Initial Size of TBW Predicts Training Related Changes.....	87
Discussion.....	88
Acknowledgements.....	91
References.....	98
<b>III. Multisensory Perceptual Learning is Dependent Upon Task Difficulty</b> .....	<b>102</b>
Abstract.....	102
Introduction.....	102
Methods.....	106
<i>Subjects</i> .....	106
<i>Stimuli</i> .....	106
<i>Experimental Procedure</i> .....	107
<i>Experimental Protocol</i> .....	107
<i>Perceptual Training</i> .....	108
<i>Grand Mean SOA Analysis</i> .....	109
<i>TBW Analysis</i> .....	109
Results.....	110
<i>Perceptual Training with Easy Training Stimuli Increases Report of</i> <i>Simultaneity while Difficult Training Stimuli Decrease Report of</i> <i>Simultaneity</i> .....	110
<i>Effect of Training Difficulty on TBW</i> .....	111
<i>Greater Enhancement of Temporal Acuity with More Difficult Training</i> <i>Protocols</i> .....	112
Discussion.....	112
Acknowledgements.....	115
References.....	119

<b>IV. The Impact of Feedback on the Different Time Courses of Multisensory Temporal Recalibration .....</b>	<b>123</b>
Abstract.....	123
Background.....	124
Methods .....	127
<i>Participants.....</i>	127
<i>Assessment of Temporal Acuity by Simultaneity Judgment Task.....</i>	127
<i>General Experimental Procedure for the Assessment of Temporal Recalibration .....</i>	128
<i>Presentation of Feedback Signal .....</i>	129
<i>Analysis of the Time Course of Multisensory Temporal Recalibration .....</i>	129
Results and Discussion.....	131
<i>Feedback Accelerates and Maintains Cumulative Temporal Recalibration ....</i>	131
<i>Positive and Negative Feedback Differentially Impact the Time Course of Temporal Recalibration .....</i>	134
<i>Time Course of Rapid Recalibration of the TBW Diverges as a Function of Prior Feedback Reliability.....</i>	136
General Discussion.....	138
Conclusions .....	141
Acknowledgements.....	141
References.....	148
<b>V. Perceptual Training Enhances Temporal Acuity for Multisensory Speech .....</b>	<b>153</b>
Abstract.....	153
Introduction .....	154
Methods .....	158
<i>Subjects.....</i>	158

<i>Stimuli</i> .....	158
<i>Simultaneity Judgment Tasks</i> .....	159
<i>Perceptual Training Sessions</i> .....	160
<i>Protocol</i> .....	161
<i>Analysis</i> .....	162
Results .....	162
<i>Perceptual Training Enhances Temporal Acuity for “Flash-beep” and</i> <i>Audiovisual Speech Stimuli</i> .....	163
<i>Changes in Temporal Acuity Remain Stable for “Flash-beep” Stimuli but not</i> <i>Audiovisual Speech Stimuli at Least 1 Week Following Perceptual</i> <i>Training</i> .....	164
<i>Perceptual Training does not Generalize Across Levels of Stimulus</i> <i>Complexity</i> .....	165
Discussion.....	165
References.....	176
<b>VI. The Neural Correlates of Enhanced Temporal Acuity for Audiovisual Speech</b> <b>Following Multisensory Perceptual Training</b> .....	<b>181</b>
Abstract.....	181
Introduction .....	181
Methods .....	184
<i>Subjects</i> .....	184
<i>Experimental Procedure</i> .....	184
Multisensory Perceptual Training Protocol .....	184
<i>Stimuli</i> .....	184
<i>SJ Assessments</i> .....	185
<i>Perceptual Training</i> .....	185

<i>Behavioral Analysis</i> .....	186
fMRI Neuroimaging Protocol.....	187
<i>Block-design localizer</i> .....	187
<i>Simultaneity Judgment Task</i> .....	188
<i>MRI and fMRI Analysis</i> .....	188
<i>Whole-brain analysis</i> .....	189
Results .....	190
<i>Perceptual Training Enhances Temporal Acuity for Audiovisual Speech</i> .....	190
<i>Report of Synchrony for fMRI SJ Assessments is Unaffected by Perceptual</i> <i>Training</i> .....	191
Whole Brain Neuroimaging Analysis.....	191
<i>Main Effects of Training Status and SOA</i> .....	191
<i>Interaction of Training Status and SOA</i> .....	191
<i>Simple Effects of Training Status at Each SOA</i> .....	192
Discussion.....	192
References.....	211
<b>VII. Discussion.</b> .....	<b>215</b>
Summary of Findings .....	215
<i>Multisensory Perceptual Learning is Dependent on Feedback</i> .....	215
<i>Multisensory Perceptual Training Enhances Temporal Acuity for Audiovisual</i> <i>Speech</i> .....	216
<i>Perceptual Training with Audiovisual Speech Stimuli Elicits Changes in a</i> <i>Wider Network of Multisensory and Decision-making Areas</i> .....	217
<i>Feedback Influences the Time Course of Cumulative and Rapid Multisensory</i> <i>Temporal Recalibration</i> .....	217
Relevant Themes Regarding Multisensory Temporal Perception and Plasticity .....	218

<i>Interaction of Top-down and Bottom-up Processing</i> .....	218
<i>Integrating the Principles of Multisensory Integration</i> .....	220
<i>Generalizability of Multisensory Perceptual Training</i> .....	226
<i>Perceptual Training that Enhances Temporal Acuity for Multisensory Speech is Less Durable</i> .....	229
Future Directions.....	231
<i>Design of Multisensory Perceptual Training Paradigm</i> .....	231
<i>Potential for Perceptual Training to Enhance Temporal Acuity for those with ASD</i> .....	233
References .....	235

## List of Figures

	Page
Figure 1-1. The Temporal Principle of Multisensory Integration .....	54
Figure 2-1. Study Methods .....	92
Figure 2-2. Only Reliable Feedback Alters the Probability of Synchrony Judgment .....	93
Figure 2-3. Decreasing Feedback Reliability Impairs Multisensory Perceptual Learning .....	94
Figure 2-4. Probability of Simultaneity Judgment Only Altered When Feedback is Present .....	95
Figure 2-5. Multiple Types of Feedback Elicit Multisensory Perceptual Learning .....	96
Figure 2-6. Temporal Acuity Prior to Perceptual Training Predicts Training Related Changes .....	97
Figure 3-1. SJ Task and Study Methods .....	116
Figure 3-2. Effect of Task Difficulty on the Report of Synchrony at each SOA.....	117
Figure 3-3. Change in TBW for Each Group.....	118
Figure 4-1. Measurement of Cumulative and Rapid Temporal Recalibration.....	142
Figure 4-2. Cumulative Temporal Recalibration .....	143
Figure 4-3. Effect of Prior Feedback on Temporal Recalibration.....	144
Figure 4-4. Effect of False Feedback on the Time Course of Temporal Recalibration.....	145
Figure 4-5. Rapid Temporal Recalibration.....	146
Figure 4-6. The Effect of False Feedback on Cumulative and Rapid Temporal Recalibration.....	147
Figure 5-1. Study Methods .....	170
Figure 5-2. Perceptual Training Influences Temporal Acuity for Audiovisual Speech .....	171
Figure 5-3. Perceptual Training Influences Temporal Acuity for “Flash-beep” Stimuli .....	172
Figure 5-4. Practice Without Feedback does not Affect Temporal Acuity for Audiovisual Speech.....	173

Figure 5-5. Perceptual Training Enhances Temporal Acuity .....	174
Figure 5-6. Perceptual Training Does Not Generalize Across Stimulus Complexity .....	175
Figure 6-1. Perceptual Training Enhances Temporal Acuity for Audiovisual Speech .....	195
Figure 6-2. Whole brain analysis of the main effects of training status and SOA.....	196
Figure 6-3. Whole brain analysis of interaction of SOA and training status.....	197
Figure 6-4. Simple main effect of training status at each SOA.....	198

## List of Tables

	Page
Table 6-1. Effects of Training Status and SOA by Center of Mass.....	199
Table 6-2. Effects of Training Status and SOA by Peak Voxel.....	200
Table 6-3. Interaction of Training Status and SOA by Center of Mass.....	201
Table 6-4. Interaction of Training Status and SOA by Peak Voxel.....	202
Table 6-5. Simple Main Effects Training Status at each SOA by Center of Mass.....	203
Table 6-6. Simple Main Effects Training Status at each SOA by Peak Voxel.....	205
Table 6-7. Effects Training Status and SOA by Center of Mass and Defined by Author.....	207
Table 6-8. Interaction of Training Status and SOA by Center of Mass and Defined by Author.....	208
Table 6-9. Simple Main Effects Training Status at each SOA by Center of Mass and Defined by Author.....	209

## Chapter I

### Introduction

The sensory signals we continuously encounter in the environment inform our understanding of the world and appreciably influence our behavior. Although the study of sensory systems has historically viewed these sensory signals in isolation (Ryan A Stevenson, Ghose, et al., 2014), an organism rarely encounters stimuli in its environment that affect only a single sensory modality. Whether it's hearing a familiar voice that helps us find a face in a crowd (Zweig, Suzuki, & Grabowecky, 2015) or how something might taste is changed by what we also see and smell (Spence, 2015), multisensory interactions enhance our perception of the world. Subsequently, there has been a growing interest in determining how this interaction of the senses guides behavior as well as manifests at the neural level (Stein, Stanford, & Rowland, 2014; Wallace & Stevenson, 2014).

The following introduction will define and describe the guiding principles of multisensory processing (with a detailed description of the temporal aspects of multisensory processing), discuss the temporal principle of multisensory integration and the perception of multisensory temporal relationships, detail concepts and theories regarding multisensory plasticity, and review general principles of perceptual learning and describe the emerging literature regarding multisensory perceptual learning. The subsequent chapters will offer new evidence regarding the manner and contingencies with which multisensory perceptual learning occurs and the capacity for multisensory perceptual learning to enhance temporal acuity for audiovisual speech. Finally, a discussion of the novel findings reported here will be contextualized in regards to the existing understanding of multisensory processing and perceptual learning as well as provide ideas for future research in these areas.

## **An Introduction to Multisensory Processing**

### *Defining Multisensory Interactions*

Although a definition of what constitutes multisensory interactions may seem intuitive, a wide variety of terms and definitions have been applied to these sensory interactions in differing ways. Particularly problematic to the multisensory field is the occasional use of the term multisensory integration to describe all manners of multisensory processes, distinct from the definition of multisensory integration employed by neurophysiologists to refer to a change in neural activity derived from the interaction of separate sensory signals (Stein et al., 2010). For a single neuron, the neuron may be considered multisensory if a response is evoked by more than one sensory modality or if the presence of an additional sensory modality alters the evoked unisensory response to a different sensory modality. Yet not all multisensory neurons may have the capacity for multisensory integration (Wallace & Stein, 1997). For a neuron, multisensory integration refers to a change in the neuron's response (typically quantified by its firing rate) that is statistically significantly different from the response evoked by the unisensory stimuli (typically in reference to the response of the most effective unisensory stimulus) (Stein & Stanford, 2008; Stein, Stanford, Ramachandran, Perrault Jr, & Rowland, 2009). Although this definition of multisensory integration may describe the processes that underlie multisensory integration in neurons, it is not easily applied to studies of behavior.

To describe multisensory interactions that may be observed at both the neural and behavioral levels, multisensory integration has been defined as “the process by which inputs from two or more senses are combined to form a product that is distinct from, and thus cannot be easily ‘deconstructed’ to reconstitute the components from which it is created” (Stein et al., 2014). As not all multisensory interactions produce distinct products that are readily observable, the term multisensory processing has subsequently been adopted to refer any process involving more than one sensory modality, independent of the interaction of the sensory cues.

In addition to these semantic distinctions, it is important to note that a number of other terms have been utilized as synonyms for the term multisensory such as cross-modal, multimodal, or polysensory. While the use of some of these terms (polysensory or multimodal) has diminished, the term “multisensory” is favored for describing more general behavioral or neural processes that arise from sensory interactions whereas the term “cross-modal” has been subsequently defined as referring to the properties of a stimulus whose features affect more than one sensory modality. Additionally, the terms amodal or supramodal have been utilized to refer to stimulus features that are not specific to a single sensory modality (e.g. the semantic relationship of the stimuli) (Stein et al., 2010).

### *The Principles of Multisensory Processing*

Multisensory processing appears to follow three primary principles that are evident at both neural and behavioral levels of study. The “temporal principle” states that the greatest multisensory interactions are evident as sensory cues from separate modalities are increasingly proximate in time. Similarly, the “spatial principle” states that the greatest multisensory interactions will occur when sensory cues are increasingly proximate in space. Collectively, these first two principles convey the concept that stimuli that are close in space and time are likely derived from the same event (Stein et al., 2014). At the neural level, these principles are evident as stimuli that are increasingly proximate in space (Meredith & Stein, 1986a; Wallace, Meredith, & Stein, 1992) or time (Meredith, Nemitz, & Stein, 1987; Stein & Wallace, 1996) exhibit response enhancement (as well as response depression for stimuli that are significantly spatially or temporally disparate). Behaviorally, these principles are also apparent, as stimuli that are temporally and spatially proximate are likely to be perceptually bound (Ryan A.

Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012; Wallace, Roberson, et al., 2004).<sup>1</sup> The third principle of multisensory processing is the “principle of inverse effectiveness” which states that the magnitude of multisensory interactions (typically those that elicit multisensory integration) is greatest when the individual sensory cues are weakly effective. At the neural level, this principle is readily observed (as well as easily measured) when separate unisensory stimuli, which may each evoke minimal neural activity, are presented simultaneously, evoking a significant increase in neural activity; however, as the effectiveness of the unisensory stimuli increases, the magnitude of multisensory enhancement decrease when the multisensory stimulus pair is presented (Meredith & Stein, 1983, 1986b). This principle is also evident in behavioral findings as, for example, the magnitude of multisensory enhancement of speech recognition increases with increasing auditory noise as the magnitude of the speech signal remains the same (although this enhancement is not observed when auditory noise is significantly increased with a speech signal of constant magnitude) (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007).

In addition to these primary principles, some have additionally posited that a fourth principle, the semantic relatedness of the stimuli, may also describe multisensory interactions (Doehrmann & Naumer, 2008; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). The primary three principles of multisensory integration are hypothesized to describe bottom-up multisensory interactions at lower-level sensory areas. In contrast, semantic relatedness describes a higher-order cognitive process for which learned associations have developed

---

<sup>1</sup> Recent evidence suggests that temporal and spatial integration are separate processes rather than a single, common mechanism governing sensory integration as one’s probability of integration for spatial stimuli does not relate to one’s probability of integration for temporal stimuli (Odegaard & Shams, 2016).

between stimuli from separate sensory modalities (Doehrmann & Naumer, 2008; Tsilionis & Vatakis, 2016). Indeed, it has been demonstrated that statistical learning can produce multisensory associations between specific auditory and visual stimuli following exposure to regular relationships between these stimuli, suggesting that cross-modal associations arise from statistical regularities in the environment and that arbitrary semantic relationships may develop from statistical learning (Aaron R Seitz, Kim, van Wassenhove, & Shams, 2007). Thus, the relatedness of multisensory stimuli may reflect the expectation for which the stimulus in one sensory modality is congruent or matches with a stimulus in a different sensory modality. For example, the congruency of the visual and auditory components of speech in terms of the gender of the speaker may influence the perception of synchrony (Vatakis & Spence, 2007). Additionally, other stimulus features may influence multisensory interactions as synesthetically congruent stimuli (e.g. low-frequency sound with larger visual disk) speeded response times for a visual size discrimination task (Gallace & Spence, 2006).

#### *Influence of Multisensory Interactions on Behavior*

As the information provided by sensory signals is essential for an organism to navigate within its environment, the capacity to integrate multiple sensory cues enhances perception and allows for more accurate perceptual evaluations of the environment. Increasing evidence suggests that auditory and visual information (as well sensory signals from other sensory modalities) are integrated in an optimal fashion to guide behavior based on information weighted from these multiple sensory modalities (Kayser & Shams, 2015; Shams, Ma, & Beierholm, 2005; Stein et al., 2014). Thus, by having additional sensory information that describe different aspects of the same event, a more accurate representation may be inferred, particularly if the sensory signal for one modality is less reliable. These concepts have further been extended by predictive coding models of multisensory processing to incorporate different stages of the sensory processing and decisional hierarchy (Olasagasti, Bouton, & Giraud,

2015). Correspondingly, these multisensory interactions have been observed to significantly affect behavior, often resulting in behavioral benefits.

Multisensory interactions produce a number of behavioral benefits including more accurate localization (Nidiffer, Stevenson, Fister, Barnett, & Wallace, 2016), enhanced perceptual learning (Aaron R Seitz, Kim, & Shams, 2006), improved visual search (Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010), better estimation of the number of events (Philippi, van Erp, & Werkhoven, 2008), and faster reaction times (Diederich & Colonius, 2004; Hershenson, 1962) than predicted by the reaction time to the fastest individual signal alone (J. Miller, 1982). Some of these behavioral benefits arise from nonspecific interactions of sensory signals that are behaviorally relevant and irrelevant. An example of a benefit of nonspecific audiovisual interaction is the “pip and pop” effect. A spatially nonspecific auditory signal may increase the salience of a synchronous visual signal in a dynamic environment that is not due to general alerting or top-down cuing of visual change but rather emerges from the integration of these sensory cues which enhances attention, possibly for visual signals (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). Furthermore, the benefits of multisensory interactions are also evident for more socially relevant behaviors as audiovisual speech perception (Sumby & Pollack, 1954) and the identification of individual emotions (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007). For audiovisual speech, visual speech cues enhance speech comprehension (Arnold & Hill, 2001) and enhance speech intelligibility in a noisy environment (MacLeod & Summerfield, 1987; Ross et al., 2007; Sumby & Pollack, 1954). Similarly, when audiovisual cues that convey emotion (i.e. visual facial expression, auditory speech tone) are present, emotional classification is superior than when the unisensory cue is only presented (Kreifelts et al., 2007).

Multisensory interactions may also result in the perception of illusory percepts that likely result from the integration of multiple sensory cues, providing further behavioral evidence of multisensory integration. One of the most extensively studied multisensory illusions is known as

the McGurk effect. The McGurk illusion arises from conflicting speech when the auditory signal for /ba/ or /pa/ are dubbed with the visual signal for /ga/ or /ta/ resulting in the perception of /da/ and /ta/ respectively (McGurk & MacDonald, 1976). Another commonly described multisensory illusion is known as the sound-induced flash illusion (SIFI). The SIFI illusion occurs when an individual presented with two “beeps” and a single “flash” in succession reports the illusion of seeing two visual “flashes” (Shams, Kamitani, & Shimojo, 2000). Collectively, these illusions provide evidence of the powerful influence of multisensory interactions on perception.

## **The Temporal Aspects of Multisensory Processing**

### *Defining and Measuring Temporal Perception*

As the plasticity of multisensory temporal acuity is a primary focus of this text, it is important to describe and review the tasks used to measure temporal perception as well as define temporal acuity. The specific examples in the following description will often reference audiovisual stimuli as the perception of audiovisual stimuli is the focus of subsequent chapters.

The temporal relationships of sensory cues are an essential amodal property for determining if two or more sensory stimuli should be perceptually bound (See Figure 1-1). To achieve perceptual coherence, the brain must resolve differences in the physical and neural propagation times of stimuli from separate modalities in order to determine which sensory stimuli should be perceptually bound. Thus, for an audiovisual stimulus, the visual stimulus will travel much faster (speed of light approximately 300,000,000 m/s) than the auditory stimulus (speed of sound at sea level approximately 340 m/s), yet upon being detected by the nervous system, the auditory signal will reach the primary sensory cortex much faster than the visual signal (Vroomen & Keetels, 2010). As such, a flexible interval of time exists during which multisensory interactions are most likely to occur. This epoch of time, in which temporally proximate stimuli are likely to be perceived as related to the same event or perceptually bound, has been termed the temporal binding window (TBW) (McGovern, Roudaia, Newell, & Roach,

2016).<sup>2</sup> The TBW is often asymmetric with regard to true synchrony (i.e. an SOA of 0 ms), such that individuals are more likely to perceive stimuli with a greater visual-lead relative in comparison to stimuli with an auditory-lead (Conrey & Pisoni, 2006; Dixon & Spitz, 1980; van Wassenhove et al., 2007). The asymmetry observed for the TBW is thought to reflect the statistics of the environment (i.e. visual stimuli precede auditory stimuli) (R. A. Stevenson & Wallace, 2013). Additionally, the point at which two signals are most likely to be perceived as simultaneous, termed the point of subjective simultaneity (PSS), is hypothesized to further reflect the environmental statistics of the world as for audiovisual stimuli, the PSS is often observed at a stimulus onset asynchrony (SOA) for which the visual stimulus precedes the auditory stimulus (Vroomen & Keetels, 2010). Both the PSS and TBW can be estimated experimentally as have been often utilized as measures to describe temporal perception. The following paragraphs provide a detailed overview of the experimental techniques often utilized to measure the temporal perception of multisensory stimuli.

A variety of tasks have been employed to measure the perception of synchrony. One of the most commonly used tasks is known as a simultaneity judgment (SJ) task in which a participant is typically asked to judge whether or not the presentation of two stimuli was simultaneous (Vroomen & Keetels, 2010). Although there are a number variants of the SJ task, the most common variant requires a participant to respond whether the two stimuli were presented synchronously or asynchronously, which is often referred to as a two-alternative forced choice (2-AFC) SJ task (however, there is some debate as to whether these categories

---

<sup>2</sup> As an important note, various synonyms have been utilized to refer to the TBW such as the window of temporal integration (Vroomen & Keetels, 2010), simultaneity window (Noel, Łukowska, Wallace, & Serino, 2016), temporal window of integration (van Wassenhove, Grant, & Poeppel, 2007), temporal integration window (Petrini, Russell, & Pollick, 2009), or simply, the window of integration (Colonius & Diederich, 2004).

[i.e. synchrony or asynchrony] are truly separate and thus this 2-AFC task may simply be more accurately described as a yes/no task as behavior data is very similar when participants are asked to respond whether or not the stimuli were simultaneous; as much of the prior literature has used the 2-AFC distinction to refer to this variant of the SJ task, this SJ task will often be referred to here after as a 2-AFC SJ task). For stimuli presented at auditory- to visual-leading SOAs (generally, auditory-leading values are denoted as negative and plotted on the left of synchrony while visual-leading values are denoted as positive and plotted on the right of synchrony), the participants' report of synchrony at each SOA typically yields a distribution resembling a Gaussian curve (although this approximation does not best represent the data as will be described in more detail). The peak of this distribution reflects the PSS while the width of this distribution may be utilized to estimate the TBW.

Complementary to the SJ task, another commonly employed task, known as a temporal order judgment (TOJ) task, asks participants to judge which of the two stimuli presented at various SOAs came first (i.e. did the auditory or visual stimulus occur first). If the probability of one stimulus category being reported as having occurred first, ex. visual stimulus occurred first, is plotted as a function of auditory- to visual-leading SOAs, the data usually resembles a sigmoid shaped curve. From this distribution, the PSS is reflected by the SOA at which a participant equally responded that either stimulus occurred first, for example the SOA at which 50% of visual stimuli were reported to occur first. Another measure, the just noticeable difference (JND) represents the smallest interval for which a participant can reliably perceive the temporal order of the stimuli (generally measured as half of the interval of time from the SOA at which a stimulus is perceived as the first stimulus on 25% of trials to the SOA at which a stimulus is perceived as the first stimulus on 75% of trials).

A variety of different data fitting methods have been employed to estimate the PSS, TBW, and JND (García-Pérez & Alcalá-Quintana, 2012). For SJ tasks, two commonly employed fitting methods, either a single Gaussian function (Eg & Behne, 2015; Fujisaki, Shimojo,

Kashino, & Nishida, 2004; Zampini, Guest, Shore, & Spence, 2005) or two separate logistic functions (Noel et al., 2016; Powers, Hillock, & Wallace, 2009; R. A. Stevenson & Wallace, 2013), are often applied to the experimentally derived data (although additional alternative fitting methods have also been utilized (van Wassenhove et al., 2007)). While the PSS is often estimated by the peak of the Gaussian curve or point of intersection of the separate sigmoid curves, a variety of criteria have been employed to measure the TBW. For Gaussian curve fits, the standard deviation (or twice the standard deviation) is often used to estimate the TBW (Zampini et al., 2005), although some have estimated the TBW by measuring the full width at half maximum of the Gaussian curve (Eg & Behne, 2015; Petrini, Russell, et al., 2009). When separate sigmoid curve fits are applied to the data, the TBW is estimated as the interval from the SOA at which a participant's probability of reporting synchrony exceeded a criterion (often a report of synchrony on 75% of trials) to the SOA at which the PSS was estimated. Unlike the Gaussian curve fit, the separate sigmoid fits allow for separate estimates of auditory- and visual-leading TBWs (as well as the whole TBW). As some have argued that audiovisual stimuli that are perceived auditory- and visual-leading stimuli should be estimated with separate psychometric functions, the double sigmoid curve fit is suggested to be advantageous as it allows for better representation of the data. For TOJ tasks, a logistic function is commonly employed to fit the experimentally derived data (Keetels & Vroomen, 2007; Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence, 2008). In contrast to the SJ task, estimates of the PSS and JND are generally obtained using the point at which 50% of one stimulus is perceived to occur and the interval from which 25% to 75% of a stimuli are perceived first for each measure respectively described above (although see (R. A. Stevenson & Wallace, 2013) for estimates of the TBW on a TOJ task as the interval from the SOA at which a stimulus was perceived first on 35% of trials to the SOA at which the same stimulus was perceived first on 65% of trials).

It is important to note that one of the disadvantages of using a Gaussian function or logistic function to fit data derived from a SJ or TOJ task is that the experimentally derived

estimates of synchrony or temporal order perception are typically not a smooth function and often the range of SOA for which synchrony is reported is typically asymmetrical (and not well described by a Gaussian function in particular) (Alcalá-Quintana & García-Pérez, 2013). Subsequently, there has been efforts to generate models that can better describe the experimentally derived data for TOJ and SJ tasks that employ various parameters to describe the factors that determine the temporal judgment made by participants performing these tasks (Alcalá-Quintana & García-Pérez, 2013; García-Pérez & Alcalá-Quintana, 2012). Furthermore, models using a causal inference modeling framework have been applied to the SJ task (Magnotti, Ma, & Beauchamp, 2013). Additional models have also sought to relate and extend the TBW to other, more general psychophysical measurements such as reaction times (Diederich & Colonius, 2015). As these models are relatively new, future research will be necessary to understand how the estimates of temporal perception relate to more classic methods of deriving measures of temporal perception.

While SJ and TOJ tasks have been extensively used to describe temporal perception, there are many considerations as to what aspects of these tasks are similar or differ. The TBW and JND derived from SJ and TOJ tasks respectively have been utilized as measures to index individuals' temporal acuity. Broadly defined, temporal acuity has been used to refer to minimum time at which one can perceive two events as nonsimultaneous or successive whereas judgments of temporal order require two events to be perceived as successive and then subsequently judged for the temporal order of the stimuli (Hirsh & Sherrick Jr, 1961). Yet while the TBW and JND may seem to measure similar processes, differences in criterion that underlie what is simultaneous and what is successive have been hypothesized to produce differences between the estimates of the TBW and JND (as well as the PSS) measured by these tasks. A comparison of 2-AFC SJ and TOJ tasks by Stevenson & Wallace (2013) provided a comprehensive comparison of the JND and TBW derived by these tasks. In this study, individual estimates of the auditory-leading TBW (with a corresponding criterion of participants reporting

synchrony on 50% of trials for the SJ task and a corresponding interval from 25% to 50% of visual stimuli perceived first, i.e. the JND) show some level of within-subjects correlation across the 2-AFC SJ and TOJ tasks for identical stimuli as well as stimuli of varying complexities; however, no significant within-subjects correlation of estimates of both the more ecologically relevant visual-leading TBW and whole TBW derived by the 2-AFC and TOJ tasks was observed between tasks for identical stimuli or non-identical stimuli. Interestingly, when a criterion of participants report of synchrony 70% of trials for the SJ task and an interval from 35% to 65% of visual trials reported first for the TOJ task were used to define the whole TBW, a greater level of within-subjects correlation was seen for whole TBWs derived from TOJ and SJ tasks, however little correlation was observed for the visual-leading TBW (R. A. Stevenson & Wallace, 2013). A similar study, which measured the TBW as the standard deviation of a Gaussian probability density function and Gaussian cumulative distribution function for SJ and TOJ tasks respectively, also reported that there was no correlation observed for the TBW across TOJ and SJ tasks for multiple audiovisual stimuli (Love, Petrini, Cheng, & Pollick, 2013). These findings suggest that comparisons of the JND and TBW derived from TOJ and SJ tasks between studies may not be informative or should be cautiously interpreted.

Similarly, the PSS derived from SJ and TOJ tasks may not be equivalent. In a comparison of the PSS measured by SJ and TOJ tasks, a difference in the average PSS reported by the SJ and TOJ tasks existed only for the more complex and dynamic stimulus, a bouncing ball, in contrast to simple stimuli (i.e. the stimuli were “flash-beep” like stimuli). Yet estimates of the PSS derived from TOJ and SJ tasks for did not correlate for either the “flash-beep” like stimuli or the bouncing ball stimuli (Van Eijk, Kohlrausch, Juola, & van de Par, 2008). Additionally, a review of studies using SJ and TOJ tasks reported that estimates of the PSS derived from TOJ tasks are more likely than SJ tasks to report an auditory-leading PSS (Van Eijk et al., 2008). It has been suggested that TOJ tasks may be more variable and provide less reliable estimates of temporal perception as participants are not presented with an option of

reporting either synchrony or that they were guessing in the range of SOAs used to measure the JND (García-Pérez & Alcalá-Quintana, 2012). Interestingly, some have suggested that a hybrid of the TOJ and SJ task, the 3-AFC SJ task, is superior to a TOJ task given the shortcomings of the TOJ task (Ulrich, 1987; Van Eijk et al., 2008). The 3-AFC SJ task presents participants with the options of responding whether the auditory stimulus occurred first, the visual stimulus occurred first, or whether the auditory and visual stimulus occurred simultaneously. The advantage of this task is that it is hypothesized to reduce the response bias introduced by the 2-AFC SJ and has been suggested as being used in place of TOJ or SJ tasks (Van Eijk et al., 2008).

Another variant of the SJ task, the 2-interval forced choice (2-IFC) SJ task, has also been employed to measure temporal acuity (Powers et al., 2009). In the 2-IFC SJ task, two pairs of stimuli separated by a temporal interval are presented on each trial (typically one of the stimulus pairs is the veridical presentation of the simultaneous stimuli, i.e. an SOA of 0 ms, while the other stimulus pair is presented at an asynchronous SOA). Participants completing a 2-IFC SJ task are typically asked to respond if the stimuli presented in the first interval or second interval were simultaneous. An advantage of the 2-IFC SJ task is that unlike the 2-AFC SJ or TOJ tasks, it is not influenced by criterion of simultaneity or successiveness. While the 2-IFC SJ task may not be affected by the same potential for changing criterion as a 2-AFC (yes/no) SJ task or a TOJ task, the 2-IFC task may still be affected by an interval bias in which participants have a response bias to report one of the response choices more frequently for one of the stimulus presentation intervals (García-Pérez & Alcalá-Quintana, 2011). Such an interval bias was observed by Powers *et al.* (2009), who reported that individuals were more likely to report stimuli presented in the second interval as simultaneous in comparison to stimuli presented in the first interval on catch trials for which both stimulus pairs were indeed simultaneous.

The subsequent chapters will describe various experiments that employed the 2-AFC SJ task. These experiments will also use of both Gaussian and double sigmoid curve fitting methods, which were determined by the type of fitting methods employed by similar studies.

### *The Relationship of Temporal Proximity, Synchrony Perception, and Multisensory Integration*

The temporal principle (as described above) as well as the concept of a TBW is based, in part, on observations from single neuron recordings (Wallace & Stevenson, 2014). In these initial reports in the cat model, *Meredith et al. (1987)* reported that temporally proximate stimuli elicited the greatest multisensory enhancement of neural activity and that SC neurons exhibited a temporal range, spanning typically 400 ms or less, in which response enhancement or depression was observed for temporally disparate stimuli (primarily audiovisual stimuli) (Meredith et al., 1987).<sup>3</sup> Similarly, cortical neurons in the anterior ectosylvian sulcus (AES) of the cat also exhibit multisensory enhancement that is greatest when stimuli were temporally proximate, suggesting this principle applies across multiple brain structures (Stein & Wallace, 1996). In these studies of single neurons, a range of multisensory interactions that were significantly different from the responses elicited from unisensory stimuli can be used to infer the

---

<sup>3</sup> Interestingly, it has recently been reported that multisensory enhancement in the SC is greatest when a neuron is equally excited by either an auditory or visual stimulus, rather than preferentially activated by one of the unisensory stimuli (R. L. Miller, Pluta, Stein, & Rowland, 2015). In the context of multisensory temporal integration, it appears that the temporal tuning function of a neuron is dependent on the SOA if the effectiveness of the unisensory is balanced; however, if the effectiveness between the unisensory stimuli differs, multisensory enhancement is greater if the more effective stimulus precedes the less effective stimulus. Thus, these findings suggest that in some situations, stimulus effectiveness and temporal alignment may interact to affect multisensory enhancement (R. L. Miller et al., 2015).

range of SOAs over which multisensory integration occurs (Stein & Stanford, 2008). Conceptually, numerous behavioral findings that report multisensory interactions within a temporal range appear to parallel the findings of single neuron recordings (Wallace & Stevenson, 2014). While multisensory integration may be easily inferred for electrophysiological recordings of single neurons, inferring whether the TBW derived from various behavioral tasks (particularly in the case of synchrony judgments) is representative of the temporal range over which multisensory integration occurs is more challenging. In particular, the perception of synchrony and the temporal range over which multisensory integration occurs have been suggested as possibly separate processes with distinct underlying neural networks (L. M. Miller & D'Esposito, 2005; Ryan A Stevenson, Altieri, Kim, Pisoni, & James, 2010; Ryan A Stevenson, VanDerKlok, Pisoni, & James, 2011; Tsilonis & Vatakis, 2016). As many of the studies in the subsequent chapters examine the perceptual report of simultaneity, some of the reported differences between synchrony perception and direct measures of multisensory integration are described.

For behavioral measures, one way multisensory integration may be inferred when behavioral exceeds the expected enhancement due to statistical facilitation. For example, faster reaction times to audiovisual stimuli with smaller SOAs are observed to be faster than predicted by statistical facilitation, suggesting that significantly more multisensory integration occurs at smaller SOAs for audiovisual stimuli (Ryan A. Stevenson, Fister, et al., 2012). Other studies of detection have also selected an experimental design that would possibly allow for multisensory integration to be inferred for improvements in performance. An interesting study observed an increase in the accuracy of detecting a change in luminance using a 2-IFC task when the intensity of an uninformative sound increased in occurred with an increase in luminance at SOAs from -150 ms to 75 ms, suggesting multisensory integration of these stimuli facilitated the increase in visual detection accuracy when the auditory and visual stimuli were proximate (Andersen & Mamassian, 2008).

Another measure of multisensory integration is the perception of an illusory percept that results from the integration of separate sensory stimuli (Ryan A Stevenson, Ghose, et al., 2014). As the report of an illusory percept may directly measure multisensory integration, the perception of synchrony and the perception of a multisensory illusion may be contrasted. The effect of temporal proximity on synchrony perception and multisensory integration has been most well studied for the McGurk illusion. A study by van Wassenhove et al. (2007) sought to address whether the TBW for the perception of simultaneity (i.e. perception of synchrony of the auditory and visual speech) was equivalent to the TBW for the perception of the McGurk illusion (i.e. the range of temporal disparity for which the McGurk illusion was still perceived. In this study, both voiced (/ba/ /ga/ /da/) or voiceless (/pa/ /ka/ /ta/) illusory McGurk pairs were presented at various SOAs to separate groups of participants. Participants were asked in one experiment to report “what they hear while looking at the face” in order to measure perception of the McGurk illusion while in another experiment they were asked to report whether the same stimuli were simultaneous; additionally, participants were asked to report whether presentations of the congruent pair corresponding to the illusory percept (e.g. visual /da/ with auditory /da/) at various SOAs were simultaneous. van Wassenhove et al. reported that while the TBW for the perception of the voiced McGurk illusion (208 ms) was nearly equivalent to the TBW for the congruent /da/ stimulus (205 ms; visual /da/, auditory /da/) was perceived to be synchronous, the TBW for incongruent illusory stimulus was much smaller (159 ms). In contrast, for the voiceless McGurk stimulus (auditory /pa/ with visual /ka/ elicits the perception of /ta/), the TBW for the perception of the McGurk illusion (161 ms) was smaller than the TBW for the perceived synchrony of congruent /ta/ (205 ms) but identical to the TBW for the incongruent McGurk stimulus (161 ms) (van Wassenhove et al., 2007). Although, the estimates of the TBW for both tasks was similar, these findings do suggest that the range of SOAs over which integration is evident is smaller than range of SOAs at which synchrony is perceived. Yet despite some of the observed differences in the reported range of SOAs at which synchrony and the illusory were

perceived, this evidence suggests that the perception of simultaneity may approximate the TBW for which multisensory integration is likely to occur. Additional evidence suggest that the temporal binding of stimulus features (or alternatively the detection of asynchrony) is a separate process from multisensory integration as the range of SOAs over which the McGurk illusion is perceived is greater than range of SOAs at which individuals perceive the stimuli as synchronous such that an individual may be able to detect the auditory and visual cues to be asynchronous while also perceiving the illusory integration of the unisensory speech cues on the same trial (Soto-Faraco & Alsius, 2009). One additional caveat to note when comparing the perception of congruent audiovisual speech and the perception of the McGurk illusion is that each may activate separate neural networks (Erickson et al., 2014). Thus, the comparison of congruent speech integration and integration resulting from the perception of the McGurk illusion may not be entirely equivalent.

The SIFI also appears to be dependent upon the temporal relationships of the stimuli such that when one “beep” (the second beep being simultaneous with the “flash”) occurs approximately 100 ms either before or after the “flash” (i.e. a TBW of ~200 ms), a second illusory “flash” is often perceived, suggestive of multisensory integration (Shams, Kamitani, & Shimojo, 2002). This TBW, however, appears to be equivalent or more narrow than the TBW reported for simple “flash-beep” stimuli (Powers et al., 2009; Ryan A. Stevenson, Zemtsov, & Wallace, 2012; Zampini et al., 2005) and thus the range of SOAs for which multisensory integration results in the illusory percept, like the McGurk illusion, may be inferred to be smaller than the range of SOAs for which a single “flash-beep” is perceived to occur simultaneously. While the SIFI appears to exhibit multisensory integration when audiovisual stimuli are temporally proximate, interestingly, some evidence suggests the spatial rule does not seem to influence the SIFI percept as the illusion appears to be perceived at a similar magnitude for spatially congruent and incongruent “flashes” and “beeps” (Innes-Brown & Crewther, 2009).

Therefore, it is possible that the principles of multisensory integration may not describe all aspects of sensory integration relating to illusory multisensory percepts.

The relationship between temporal acuity and the magnitude of multisensory integration has been further illustrated by a study that explored whether temporal acuity (i.e. the perception of synchrony) could predict the magnitude of multisensory integration (i.e. the strength of an multisensory illusion). Evidence from this study suggested that temporal acuity for low-level, “flash-beep” stimuli (as indexed by the TBW derived from a SJ task) predicts individual differences in magnitude of multisensory integration (as indexed by the perception of the McGurk effect). In particular, as the size of the visual-leading TBW decreased, the probability of McGurk perception is observed to increase. Similarly, the size of the visual-leading TBW also predicts the individual susceptibility to the SIFI. Individuals with smaller visual-leading TBWs were observed to be less susceptible to perceiving the SIFI, suggesting that those individuals with better temporal acuity were better able to dissociate asynchronous auditory and visual stimuli as are presented for the SIFI (Ryan A. Stevenson, Zemtsov, et al., 2012). Collectively, this study suggests that that the perception of synchrony provides an approximation of the magnitude of multisensory integration.

### *Factors Influencing Temporal Perception*

Numerous factors may influence how temporal relationships between multisensory stimuli are perceived. The following section details some of the stimulus features that may influence multisensory temporal perception. In particular, this section will emphasize factors that affect the TBW for audiovisual stimuli.

Multisensory temporal perception is observed to be highly variable amongst individuals (Noel, Wallace, Orchard-Mills, Alais, & Van der Burg, 2015; R. A. Stevenson & Wallace, 2013; Van Eijk et al., 2008). Yet despite these this range of variability, the tendency for temporal integration remains stable over time (Odegaard & Shams, 2016). Similarly, temporal acuity (i.e.

the TBW, particularly as measured by a SJ task) for stimuli of differing complexity is correlated for an individual subject (R. A. Stevenson & Wallace, 2013). Furthermore, for an individual, the TBW is reported to be correlated across different sensory pairing (for auditory, visual, and tactile stimuli) while exhibiting variability in TBW between individuals (Noel et al., 2015).<sup>4</sup> Thus, temporal perception is highly individualistic.

One of the more notable stimulus features that influences temporal perception is the relative complexity of the stimuli. In general, the best temporal acuity (i.e. smallest TBWs) has typically been reported for simple, “flash-beep” type stimuli (R. A. Stevenson & Wallace, 2013; Zampini et al., 2005) although for other more relatively complex stimuli, such biological motion (in particular point-light walkers), have been reported to exhibit a narrow range of SOAs over which the auditory and visual cues were perceived to be synchronous (Mendonça, Santos, & López-Moliner, 2011). Generally, most studies have also reported that the TBW for speech is larger than other complex non-speech stimuli (Dixon & Spitz, 1980). A recent study that examined the effect of stimulus complexity reported that the TBW for speech stimuli (utterances of /ba/ and /ga/) was larger than TBWs for a simple “flash-beep” stimuli or non-speech biological motion (R. A. Stevenson & Wallace, 2013). Some studies have actually reported better temporal acuity for audiovisual speech (syllables) stimuli compared to musical stimuli (musical notes on a guitar or piano) (Vatakis & Spence, 2006b). A similar study also reported that while better temporal acuity for object actions in comparison to either speech (sentences) or musical stimuli (music played on a guitar or piano), better temporal acuity for speech stimuli was observed compared to musical stimuli; however, in a follow-up experiment when complexity of the stimuli was reduced (ex. syllable compared to sentence), no effect of stimulus complexity was observed. Thus, it has been suggested that an important aspect of stimulus complexity is

---

<sup>4</sup> In contrast, the PSS does not correlate across different sensory pairing (for auditory, visual, and tactile stimuli) (Noel et al., 2015).

continuously varying properties of a stimulus compared to those stimuli with less associated auditory and visual cues (Vatakis & Spence, 2006a). Some have suggested that speech may be a special type of multisensory stimulus (Vatakis, Ghazanfar, et al., 2008). To determine if the perception of speech uniquely elicits wider temporal binding windows, *Vroomen & Stekelenburg (2011)* compared temporal acuity for natural speech, sine-wave speech (for which auditory signal was reduced to a few sinusoids that followed the centre frequency and the amplitude of the first three formants) that was not perceived as speech, and sine-wave speech that was perceived as speech. In this study, the TBW and PSS measured using a SJ task were observed to be equivalent for all stimuli, suggesting that the perception of speech as in fact speech but rather reflect difference in the nature of stimulus characteristics possibly due to the correlation of the time-varying aspects of the visual and auditory cues (Vroomen & Stekelenburg, 2011).

Prior sensory experience is another factor that might interact with stimulus complexity to influence sensory perception. For example, experience appears to alter the temporal representation of speech stimuli such that the PSS is observed when the visual-lead is greater for one's native language or a familiar language than a language for which an individual has minimal experience (Navarra, Alsius, Velasco, Soto-Faraco, & Spence, 2010). Experience has also been reported to enhance temporal acuity (i.e. greater capacity to detect asynchrony) for non-speech biological motion stimuli (Petrini, Dahl, et al., 2009; Petrini, Holt, & Pollick, 2010). The effect of experience was particularly notable in one study of the perception of synchrony for biological motion stimuli (point-light drumming motions for which the motion of the arm and drum stick tip were represented as well as the drum head), which observed that novice drummers needed to view the drum stick and drum head rather than just the motion of the arm in order to reliably judge the synchrony of the of the auditory and visual stimuli, suggesting the it was necessary to view the point of impact, whereas experienced drummer were able to judge the simultaneity of the auditory and visual stimuli with the arm motion alone (Petrini, Russell, et al.,

2009). Similar to prior experience, learned relationships between separate sensory stimuli also likely affect temporal perception.

The relatedness of individual unisensory stimuli (i.e. that the stimuli go together) may influence whether the sensory signals will be perceptually bound. The “unity assumption” has been used to refer to the cognitive factors that influence whether an individual is likely to attribute separate sensory signals to the same event (Vatakis & Spence, 2007). Evidence that the semantic relatedness of sensory signals affects temporal acuity derives from a study that observed enhanced temporal acuity (smaller JNDs) when speech stimuli (words and syllables) were presented with mismatched gender (e.g. a visual signal of a female speaker presented with a male voice) than when the gender of the visual and auditory signals was matched. Furthermore, smaller JNDs were reported when the gender of the speaker was congruent, but the visual and auditory signals were incongruent in terms of the syllable spoken in comparison to the congruent visual and auditory signals (Vatakis & Spence, 2007), similar to the findings of *van Wassenhove et al. (2007)*, (described above) that the TBW for incongruent syllables was smaller than the TBW for congruent syllables (van Wassenhove et al., 2007). Some evidence suggests that the semantic congruency of the auditory and visual signals may only affect temporal acuity for speech stimuli as temporal acuity is not affected by stimulus congruency for dynamic non-speech stimuli, including musical stimuli (Vatakis & Spence, 2008). Furthermore, another study demonstrated that temporal acuity was only altered by stimulus congruency for speech stimuli, but not for monkey vocalizations or human imitations of monkey vocalization, suggesting the content of human speech utilized for a TOJ task may be specific rather than an attribute related generally to communication (Vatakis, Ghazanfar, et al., 2008). Other evidence, however, suggests that stimulus congruency does affect non-speech stimuli as larger TBWs were observed for novice drummer for congruent drum beats for which the variation between auditory and visual information was maintained in comparison to TBWs for incongruent drum beats for which the variation between auditory and visual information was scrambled (although

no difference was observed in the TBW for experienced drummers when comparing the TBWs for congruent and incongruent drum beat stimuli) (Petrini, Dahl, et al., 2009).

The spatial location of stimuli is another factor that influences temporal perception. Spatial discrepancy between auditory and visual signals is one feature that appears affect temporal acuity. When auditory and visual stimuli are presented from the same position in space relative to when these stimuli are presented at different spatial locations, JNDs and TBWs are reported to be larger as measured by TOJ (Zampini, Shore, & Spence, 2003) and SJ tasks respectively (Zampini et al., 2005). The distance from a stimulus may also influence temporal perception. Although numerous studies have examined the effect of distance on the PSS, some have reported that the PSS changes with distance while other have observed the PSS remains constant with distance (Vroomen & Keetels, 2010). Some of the differences reported in these studies may be attributable to simulated distance compared to actual stimuli presented at different distances. Temporal perception may also be influenced by reference frames, such as the perceived relationship of the stimuli to the space surrounding the body (i.e. peripersonal space). Recent evidence suggests that the TBW is larger for simple, “flash-beep” like stimuli presented within peripersonal space in comparison to the TBW for “flash-beep” like stimuli presented outside peripersonal space (Noel et al., 2016).

Other stimulus properties, such as the effectiveness of individual unisensory stimuli, may influence temporal perception. Individuals are more likely to perceive less effective audiovisual stimuli (in terms of luminance and sound intensity) as synchronous over a wider range of SOAs than more effective audiovisual stimuli for individuals asked to make a simultaneity judgment (Fister, Stevenson, Nidiffer, Barnett, & Wallace, 2016). Similarly, for audiovisual stimuli on a TOJ task a more positive PSS was reported with a low intensity visual stimulus relative to PSS for high intensity visual stimulus (although the intensity of the visual stimulus was not observed to effect the JND) (Boenke, Deliano, & Ohl, 2009). Collectively, these findings suggest that for less effective stimuli, tolerance for asynchrony increases and the PSS tends to be reported at a

more ecologically relevant value. The luminance of a visual stimulus, however, may also bias temporal order judgments as stimuli with lower luminance (i.e. the less visually intense stimulus) have been reported to bias individuals towards reporting the stimulus with lower luminance as the first stimulus to appear (Bachmann, Pöder, & Luiga, 2004).

Additional factors influencing multisensory temporal perception (although not necessarily the perception of temporal synchrony) include the rate at which the relevant stimuli occur. The rate of stimulus occurrence may be particularly important for stimuli such as speech, for which the rate at which syllables are produced is approximately 4 Hz (Vroomen & Stekelenburg, 2011). The limit of audiovisual integration has been suggested to be 4 Hz as a study observed that crossmodal matching of audiovisual stimuli does not occur when the stimuli are presented at a rate exceeding 4 Hz (Benjamins, van der Smagt, & Verstraten, 2008). Interestingly, a few studies have explored if the PSS and TBW may be dependent on the tempo or rate at which the stimuli occur using dynamic stimuli that exhibit a tempo (e.g. a drumming motion). Such initial reports indicated that for stimuli with a faster tempo, the TBW became smaller and the PSS approached true synchrony such the visual-lead time decreased. Additionally, a stimulus presentation with a fast, but variable tempo appeared to shift the PSS toward an auditory leading value whereas a slow, but variable tempo is perceived with a more ethological visual-leading PSS (Arrighi, Alais, & Burr, 2006). An extension of these findings observed that only novice drummers show less sensitivity to asynchrony (i.e. wider TBWs) at slower drumming tempos (the TBW of experienced drummers was not observed to be influenced by tempo), suggesting that individuals are more likely to perceptually bind audiovisual cues at slower presentation rates (Petrini, Dahl, et al., 2009). Additionally, other behaviors may be influenced by the presentation of audiovisual stimuli at regular rates (in comparison to randomly varying the rate of presentation), which make stimuli more predictable (Marchant & Driver, 2012).

### *Neural Correlates of Multisensory Temporal Perception*

Although many different cortical areas have been observed to respond to multisensory interactions, the STS has been identified as particularly important for the interaction of auditory, visual, and somatosensory cues (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Beauchamp, Yasar, Frye, & Ro, 2008). In particular, the STS and superior temporal gyrus (STG) have been reported as important areas for the perception of audiovisual speech (Barrós-Loscertales et al., 2013; Lee & Noppeney, 2011). Subsequently, many investigations regarding the perception of multisensory temporal relationships and the perception of synchrony have focused on changes in activity in the superior temporal cortex. In this section, some of the studies of multisensory temporal perception will be reviewed.

Many studies of temporal perception have sought to describe the differences in neural responses to synchronous and asynchronous stimuli. One of the initial studies of the neural basis of multisensory perception sought to apply criteria typically used to describe multisensory integration. In this study, superadditive responses to synchronized audiovisual stimuli (white noise burst and visual reversing checkerboard pattern) and response depression to asynchronous stimuli was observed in the SC, insula, and left STS (Calvert, Hansen, Iversen, & Brammer, 2001). It is not surprising that areas such as the SC and STS were observed to be responsive to multisensory synchrony as these areas are widely reported to be responsive to multisensory stimuli. Numerous other studies have also observed these areas to be responsive to multisensory temporal relationships, however reports of activity evoked by either synchrony or asynchrony from many studies are often in disagreement.

Synchrony has been reported to enhance activity in the STS (Marchant, Ruff, & Driver, 2012; Noesselt et al., 2007), STG, supplementary motor area, left post central gyrus, left putamen/thalamus, right insula, right caudate, (Marchant et al., 2012), bilateral lateral occipital complex (LOC), SC, extrastriate visual cortex, and a region including the fusiform (FG) extending into the cerebellum (Ryan A Stevenson et al., 2010). Less activity, however, has been

reported for synchronous stimuli in the SC, anterior insula, and anterior intraparietal sulcus (IPS) (L. M. Miller & D'Esposito, 2005). Thus, for some areas such as the SC and insula, greater and lesser activity has been observed for synchronous stimuli in comparison to asynchronous stimuli.

The conflicting reports describing changes in neural activity in response to multisensory synchrony and asynchrony may be due to separate processes for the perception of multisensory stimuli and temporal relationships. Some evidence suggests that multisensory responsive regions of superior temporal cortex distinctly respond to either multisensory stimuli or properties of multisensory stimuli, such as the perception of synchrony (Ryan A Stevenson et al., 2010). *Stevenson et al. (2011)* sought to address possible differences in areas of multisensory superior temporal cortex (mSTC) that respond to synchrony/asynchrony (more responsive to synchronous than asynchronous stimuli) and audiovisual stimuli (responsive to auditory and visual stimuli), specifically audiovisual speech. In this study, synchrony defined mSTC activity was driven by synchrony independent of perceptual fusion such that higher levels of activity were observed in response to synchrony while the audiovisual defined mSTC activity was driven by perceptual fusion independent of temporal synchrony such that unfused trials elicits more activity compared to fused trials as well as greater activity for the asynchronous stimuli compared to the synchronous stimuli (Ryan A Stevenson et al., 2011). These findings further suggest a distinction in the neural correlates of perceptual fusion and the perception of synchrony. True synchrony and other SOAs may also elicit different activity than the SOA at which the stimuli are most likely to be synchronous (i.e. the PSS). The left middle STS has been reported to exhibit more activity in response to audiovisual speech (with a typical auditory delay) than to audiovisual speech that has articulatory lip movements that begin simultaneously with the auditory speech signal (Balk et al., 2010). Thus, comparisons of activity evoked by objectively synchronous stimuli to some SOAs may not always correlate with temporal perception.

Different regions of the multisensory superior temporal sulcus complex appear to exhibit various responses to visual-leading and auditory-leading asynchronous stimuli as well as synchronous stimuli (Noesselt, Bergmann, Münte, Heinze, & Spence, 2012). Some have suggested that the temporal relationships of stimuli allow for a predictive coding mechanism to enable natural continuous signals from multiple senses to be perceived as a coherent percept. A predictive coding mechanism would hypothesize that audiovisual stimuli would induce an error signal depending on the leading stimulus such that visual leading stimuli would generate an error signal in the auditory system while auditory leading stimuli would generate an error signal in the visual system. Recent evidence supports a predictive coding mechanism is utilized to temporally bind multisensory signals as visual leading stimuli (audiovisual speech, sinewave speech, and music) has been reported to increase activity (relative to synchrony) in the auditory system from Heschl's gyri into STG/STS while auditory leading stimuli (relative to synchrony) increase activity in bilateral occipito-temporal cortices and extended into STG/STS (Lee & Noppeney, 2014). Interestingly, greater activity has been reported for synchrony compared to asynchrony when stimuli were predictable rather than unpredictable in right inferior parietal cortex (Marchant et al., 2012), however it is interesting to speculate that this is not necessarily contradictory as the anticipation of a temporal relationship may reduce error signals that would evoke greater activity to synchrony.

### **Multisensory Plasticity**

Collective work examining the development of multisensory integration suggests that three developmental stages exist such that multisensory systems are initially immature, then broadly tuned, and then progressively narrowly tuned (Murray, Lewkowicz, Amedi, & Wallace, 2016). Low-level stimulus characteristics (e.g. temporal or spatial relationships) that are reflective of the sensory environment as well as learned associations regarding multisensory stimuli interact to create perceptual coherence. Throughout the course of development and into

adulthood, stimulus characteristics and learned associations result in plastic changes, eliciting both short-term and more lasting changes in perception.

The following section will address evidence for multisensory plasticity in animal models during development and adulthood, describe behavioral findings of rapid and short-term multisensory plasticity in humans, and provide an overview of unisensory and multisensory perceptual learning. This section will also place an emphasis on temporal perception and it relates to multisensory plasticity.

### *Development, Experience, and Plasticity in Multisensory Model Neural Systems*

Animal models of multisensory processing have provided a great deal of information regarding the properties of multisensory neurons. The most extensively studied neural structure has been the superior colliculus (SC) in the cat as well as the monkey (Wallace, 2004). In the cat, multisensory SC neurons are not evident at birth. Furthermore, multisensory SC neurons in the cat are initially not capable of multisensory integration as multisensory integration is evident only later in development (Wallace & Stein, 1997). Similar to the SC, the development of the capacity for multisensory integration in the anterior ectosylvian sulcus (AES) of the cat also occurs slowly (Wallace, Carriere, Perrault, Vaughan, & Stein, 2006). In contrast to the cat, multisensory neurons are present at birth in the SC of the rhesus monkey, yet these SC neurons do not have the capacity for multisensory integration, suggesting that experience may be necessary for the development of multisensory integration (Wallace & Stein, 2001). Subsequently, numerous studies have explored the role of sensory experience in the development and plasticity of multisensory processing in these animal models.

To examine the function of sensory exposure (and more likely sensory experiences) during development, a number of studies have examined the effects of sensory deprivation during development in the cat model. Early evidence suggested that visual experience might be necessary for the development of multisensory integration in the SC as lack of multisensory

enhancement is observed for audiovisual stimuli (and surprisingly auditory-somatosensory stimuli) following the absence of visual stimuli during development (Wallace, Perrault, Hairston, & Stein, 2004). More recently, evidence suggests that experience with multisensory pairs (e.g. an audiovisual stimulus) is required for the development of multisensory integration in response to that specific sensory pairing (audiovisual stimulus pair) in SC neurons that have a general capacity for multisensory integration (e.g. show multisensory integration for auditory-somatosensory pair that was also experienced during development of multisensory integration). Thus, it is likely that specific cross-modal stimulus pairs must be experienced for integrative capacity to develop rather than a general, nonspecific integrative capacity for all sensory stimuli (Xu, Yu, Stanford, Rowland, & Stein, 2015). Evidence suggests that sensory experience is also likely to be necessary for the development of typical multisensory interactions in cortical regions. In the absence of visual experience during development, spatiotemporally coincident stimuli elicit response depression at the level of single neurons in the anterior ectosylvian sulcus (AES) of the cat whereas with sensory experience these stimulus combination during normal development with visual experience would result in response enhancement (Carriere et al., 2007).

A lack of visual experience early in development may not preclude the development of multisensory integration in early adulthood. Following visual deprivation during development (which typically prevents the development of multisensory integration for SC neurons in response to audiovisual stimuli), the development of the capacity for multisensory integration has been recently observed to occur in cat SC neurons in early adulthood following experience with spatiotemporally coincident stimuli. In this study, cats that were visually deprived during development were later exposed to audiovisual stimuli during early adulthood. Prior to visual exposure, multisensory integration was not observed. Following visual exposure, the development of multisensory integration was only observed for those animals presented with spatiotemporally coincident audiovisual stimuli in contrast to animals presented with temporally

interleaved stimuli (separate presentations of each individual unisensory stimulus) at the same spatial location or separate presentations of unisensory stimuli in a random order and at random locations in space. Interestingly, the integrative capacity of the SC neurons for cats exposed to spatiotemporally coincident stimuli in this experiment continued to increase with continued experience until MSI values were no lower than observed for normally reared adult cats. These findings suggest that SC neurons retain the capacity to develop multisensory integrative capacity following the normal period of multisensory development and that experience (particularly with spatiotemporally coincident stimuli) is necessary to elicit the development of multisensory processing in these neurons (Yu, Rowland, & Stein, 2010).

The ability for multisensory SC neurons in cat to integrate temporally proximate stimuli also appears to develop with sensory experience. The range of SOAs over which multisensory integration is observed in the SC of the cat is initially small during the course of development while for adults the range of SOAs for which multisensory integration is observed is often much larger (Wallace, 2004). Indeed, the earliest appearance of multisensory integration is observed only for stimuli presented simultaneously (at true synchrony). During the course of development, the range of SOAs for which multisensory integration is observed gradually becomes larger, approaching a range of 250 ms for adult SC neurons (Wallace & Stein, 1997). Some have speculated that the increase in the temporal range for which multisensory integration occurs during development may relate to a transition from experiences that occur primarily close to the body to more experiences with distant events (Wallace & Stein, 1997).

The importance of the temporal relatedness of separate sensory stimuli was addressed by a study that explored the development of multisensory integration in SC neurons for cats in an altered sensory environment. In this study, cats were deprived of visual stimuli during development with the exception of experimentally presented auditory and visual stimuli that were temporally coincident but presented at different spatial locations. In a subset of SC neurons, multisensory integration developed, however, these neurons responded to audiovisual

stimuli that were spatially disparate (rather than exhibiting multisensory integration to spatially proximate stimuli following development in a normal sensory environment). The results of this study suggest that there is no a priori spatial relationship between sensory stimuli that is required for the development of multisensory integration and that the temporal synchrony of sensory signals is a primary factor in the development of multisensory processing. Furthermore, the results of this study also support the hypothesis that sensory experience is highly influential on the development of multisensory processing (Wallace & Stein, 2007).

Additional evidence also suggests that early experience with spatiotemporally coincident stimuli is sufficient to maintain the capacity for multisensory integration, even following visual deprivation later in development. In another study, cats that were exposed to limited but spatiotemporally coincident audiovisual stimuli (cats were deprived of all other visual stimuli during development and early adulthood) during the first 8 months of development and subsequently deprived of visual experience for approximately 12 months were reported to exhibit a capacity for multisensory integration that resembled normal animals. In contrast, cats exposed to separate, random presentations of auditory and visual stimuli did not develop the capacity for multisensory integration. These results suggest that simple exposure to spatiotemporally coincident audiovisual stimuli is sufficient for the development of multisensory integration and that further experience is not necessary to maintain the capacity for multisensory integration. Additional findings from this study also suggest that the statistics of the environment during development influence temporal filters of SC neurons as the SC neurons of cats exposed to spatiotemporally coincident stimuli in the altered sensory environment exhibited a preference for audiovisual stimuli with an SOA of 0 ms (typically, SC neurons of cats that develop in a normal sensory exhibit a preference for SOAs with a visual-lead of 50-150 ms). Collectively, the results of the previously described studies suggest that experience with temporally coincident multisensory stimuli facilitates the development of multisensory integration, although there may not exist an a priori expectation of the temporal relationships between these sensory signals.

More recently, investigations have begun to explore multisensory plasticity in adulthood, particularly as this plasticity relates to rapid or consistent changes in statistical relationship of sensory signals. A level of plasticity to accommodate for rapid adaptations to immediate changes in the statistics of the environment might be advantageous (interestingly, rapid forms of perceptual plasticity have also been of recent interest, see below and discussion). One recent study of SC neurons in adult cats observed rapid plasticity in response to exposure to audiovisual stimuli. In this study, sequences of spatially aligned visual and auditory stimuli were presented at the minimal SOA that evoked separate responses to each sensory signal (i.e. the same audiovisual stimulus pair was repeatedly presented). Following exposure to the audiovisual stimuli, an increase in the magnitude and duration in response to the first stimulus as well as a shortening of the latency and increase of the response to second stimulus was observed such that the previously separate responses appeared to merge. These findings suggest that the response to the second became associated with the response to the first stimulus (Yu, Stein, & Rowland, 2009).

Additionally, brief but repeated multisensory exposure is observed to enhance neural responses of adult SC neurons. Following repeated exposure to an audiovisual stimuli, adult SC neurons that were responsive to both auditory and visual stimuli showed significantly enhanced responses to at least one sensory modality (although the multisensory integration index MSI for most neurons was unaltered due to proportion increases in responses). More interestingly, for many SC neurons that were capable of multisensory integration but only responsive to a single sensory modality, exposure to audiovisual stimuli (but rarely unisensory stimuli) elicited responses to the sensory modality that previously did not elicit a response for that neuron. Such suggests that repetitive exposure may elicit experience dependent plasticity in adult neurons (Yu, Rowland, Xu, & Stein, 2013).

### *Multisensory Recalibration*

As an organism continuously experiences novel or changing sensory environments, the capacity to update sensory representations might be a normal aspect of perception (Ghose, 2004). Recent evidence suggests that sensory driven plasticity continues to update and adjust cognitive representations of the environment. The term recalibration has been used to describe the adjustments in sensory representations that occur in response to sensory cues that conflict with existing sensory representations. In particular, the recalibration of the perceptual representations of the temporal relationships of sensory cues has been hypothesized as a way to accommodate variability in stimulus transduction times (Fujisaki et al., 2004) or immediate changes in the sensory environment. Although recalibration has also been described for other cognitive processes, such as speech perception, (Vroomen, van Linden, De Gelder, & Bertelson, 2007), this section will focus specifically on temporal recalibration.

Initially, multisensory temporal recalibration was thought to only occur after extensive adaptation. Following repeated exposure to simple (i.e. “flash-beep” like) audiovisual stimuli with a constant SOA, temporal recalibration of the PSS in the direction of the leading sensory signal is observed (Fujisaki et al., 2004; Vroomen, Keetels, De Gelder, & Bertelson, 2004). Here, this type of temporal recalibration will be referred to as temporal recalibration following adaptation (i.e. repeated exposure to a sensory pair at a particular SOA). Generally, the effects of temporal recalibration following adaptation are reported to be similar whether the PSS is estimated by a SJ or a TOJ task, suggesting temporal recalibration is not dependent on the type of temporal judgment (Vroomen et al., 2004). While the change in the PSS has been utilized as the primary measure of temporal recalibration, the TBW appears to also widen in the direction of the leading sensory signal following adaptation (Fujisaki et al., 2004). Interestingly, temporal recalibration following adaptation does not appear to be dependent on the co-localization of auditory and visual stimuli as temporal recalibration is observed to be of equal magnitude when audiovisual stimuli are presented from the same or different locations (Keetels & Vroomen, 2007).

More recent investigations have further informed our understanding of temporal recalibration following adaptation. The effect of temporal recalibration following adaptation does not appear to be stable as the magnitude of temporal recalibration declines following initial adaptation, even if an individual is re-exposed to the adaptation stimulus. Furthermore, the dissipation of the effect of temporal recalibration following adaptation is likely due to exposure to new SOAs (i.e. new sensory evidence) as a period of delay between adaptation and measuring the temporal perception does not mitigate the effect of temporal recalibration following adaptation (Machulla, Di Luca, Froehlich, & Ernst, 2012). Findings also suggest that temporal recalibration following adaptation is not uniform but results from a difference between the test SOA and the adaptation SOA (Roach, Heron, Whitaker, & McGraw, 2011). Although many studies have primarily examined the effect of asynchronous audiovisual stimuli on temporal recalibration, temporal recalibration has also been observed for other sensory pairings as adaptation also elicits temporal recalibration for auditory-tactile (Navarra, Soto-Faraco, & Spence, 2007) and visual-tactile stimuli (Keetels & Vroomen, 2008).

Temporal adaptation also may generalize, rather than be specific to the adapted stimulus, as adaptation with simple “flash-beep” like stimuli is reported to alter the temporal representation for the stream/bounce illusion (the presentation of two moving balls results in the perception of the balls bouncing off each other when the paths of the balls cross or the perception of the balls passing by each other when sound is absent or asynchronous with the time at which the balls’ paths cross) (Fujisaki et al., 2004). Additional evidence suggests that adaptation may elicit temporal recalibration that generalizes across levels of stimulus complexity as participants that completed a TOJ or SJ task for “flash-beep” like stimuli while simultaneously monitoring asynchronous speech were observed to have a decrease temporal acuity as measured by the JND and TBW and a shift in the PSS estimated by the SJ task (Vatakis, Navarra, Soto-Faraco, & Spence, 2008). A similar finding observed that monitoring an asynchronous musical stimulus while completing a TOJ task with “flash-beep” like stimuli also

decreased temporal acuity (as measured by the JND) but did not alter the PSS (Navarra et al., 2005). Collectively, these results suggest that the generalizability of temporal recalibration following adaptation may be dependent on the task (either a TOJ or SJ task) used to measure temporal perception. Temporal recalibration following adaptation may also generalize to affect the processing of individual sensory signals. Adaptation has been reported to influence response times to unisensory stimuli as following adaptation to visual-leading audiovisual stimuli, faster responses are observed to auditory stimuli while following adaptation to auditory-leading audiovisual stimuli, slower responses times are observed to auditory stimuli (while neither adaptation stimulus affected responses times to visual stimuli), suggestive of changes in general sensory processing (Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009).

Various ideas have been positing to account for the effects of temporal recalibration following adaptation. A recent study sought to apply different models to examine three explanations of temporal recalibration defined as the latency-shift hypothesis, which suggests a change in sensory transduction time to account for changes in the temporal relationships of the auditory and visual signals, the criterion-change hypothesis, which suggests a change at higher level decisional areas changes the evaluation of audiovisual asynchrony, and the population-code hypothesis, that neural tuning of a population of neurons adjusts to reflect the temporal asynchrony between auditory and visual signals. Interestingly, all models were observed to account for the behavior of some individuals although the latency-shift and criterion-change models best represented the behavior of the population (Yarrow, Minaei, & Arnold, 2015). Thus, the effects of temporal recalibration may describe a number of phenomena.

More recently, it has been reported that temporal recalibration also occurs at very short, rapid timescales (as small as a trial-to-trial basis). *Van der Burg et al. (2013)* first observed that the PSS was observed to shift on a trial-to-trial basis, termed rapid recalibration, dependent upon the temporal relationship of auditory and visual stimuli (but not dependent on the perception of synchrony) on the immediately preceding trial such that the PSS was shifted in the

direction of the leading sensory signal. Furthermore, it was reported that rapid recalibration occurred for stimuli presented on trial t-1 at a wider range of visual-leading SOAs compared to only very small auditory-leading SOAs (as well as large SOAs, irrespective of the direction of the leading sensory signal) similar to the typical range of the TBW (Van der Burg, Alais, & Cass, 2013). That rapid temporal recalibration also occurred whether a participant was asked to judge the simultaneity of the previous trial or whether the previous trial was presently passively suggests that rapid recalibration does not result from criterion shifts (Van der Burg et al., 2013). Rapid temporal recalibration does not only occur for simple, “flash-beep” like stimuli. More recent studies have reported that rapid temporal recalibration is also observed for more complex stimuli, such as audiovisual speech (Van der Burg & Goodbourn, 2015).

Rapid recalibration may represent a form of rapid sensory plasticity and has been observed to occur within a single sensory modality as auditory frequency modulation perception exhibits dependence up the previous trial (Alais, Orchard-Mills, & Van der Burg, 2015). Yet rapid recalibration of temporal perception appears to be a multisensory process, as rapid recalibration of temporal perception is not observed for unisensory presentations auditory or visual stimuli (Harvey, Van der Burg, & Alais, 2014). Interestingly, rapid temporal recalibration also appears to be specific to audiovisual stimulus pairs as rapid temporal recalibration is not observed for audiotactile or visuotactile stimulus combinations (Van der Burg, Orchard-Mills, & Alais, 2015) unlike temporal recalibration following adaptation for auditory-tactile (Navarra et al., 2007) and visual-tactile stimuli (Keetels & Vroomen, 2008).

Some recent evidence (as well as evidence provided in Chapter 4) suggests that rapid temporal recalibration and temporal recalibration following adaptation are separate processes and may occur simultaneously but at different timescales. *Van der Burg et al. (2015)* demonstrated that when participants are asked to make simultaneity judgments following a period of adaptation, the effects of adaptation decay after the adaptation period ends while rapid temporal recalibration (i.e. trial-to-trial changes in the PSS) occurs continuously and is only

dependent upon the temporal relationship of the stimuli on the previous trial (Van der Burg, Alais, & Cass, 2015). Rapid temporal recalibration is also observed for more complex stimuli, such as audiovisual speech. Furthermore, the rapid temporal recalibration of audiovisual speech is not dependent on whether the speaker the each trial and the previous trial were the same speaker or whether the gender of the auditory and visual speech signals were congruent, suggesting that sensory features elicit rapid temporal recalibration (Van der Burg & Goodbourn, 2015).

Some questions remain as to whether temporal recalibration is a form of momentary plasticity or whether it should be considered a form of perceptual learning. Temporal recalibration that follows adaptation, as mentioned above, diminishes shortly after novel sensory experience and is not likely to elicit lasting changes in perception (Machulla et al., 2012). Yet, it could be argued that as the effect of temporal recalibration following adaptation is maintained during a delay period, these changes are quite lasting, but continuously updated and improved if new sensory evidence becomes available. Similarly, rapid temporal recalibration continuously incorporates new sensory information to alter temporal representations. It is possible that the cumulative effect of rapid temporal recalibration may, over time, facilitate a more durable form of temporal recalibration.

### *An Introduction to Perceptual Learning*

Perceptual learning is defined as relatively long-lasting changes in perception that improve an organism's ability to respond to its environment (Goldstone, 1998). The changes in perception that result from perceptual learning are driven by the environment and may not necessarily be contingent upon feedback or reward (perceptual learning is observed following both sensory exposure and sensory training (A. Seitz & Watanabe, 2005)). Unlike sensitization, habituation, or priming, which result in transient changes in perception, perceptual learning is believed to produce more lasting changes in perception (Manfred Fahle, 2005). Some forms of

perceptual learning may even be evident years later after learning (Karni & Sagi, 1993). Although perceptual learning has been studied for multiple sensory modalities, visual perceptual learning has been most extensively studied form of perceptual learning (Aaron R Seitz & Dinse, 2007). Until recently, the majority of studies of perceptual learning focused on either visual perceptual learning or other forms of unisensory perceptual learning. Yet as perceptual learning is driven by interactions with the environment, the study of unisensory perceptual learning does not likely reflect perceptual learning in more realistic settings (Shams & Seitz, 2008). Thus, similar to the development of the study of multisensory processing, there has been an increasing interest in exploring multisensory perceptual learning (Shams & Seitz, 2008). The following sections will address different aspects of perceptual learning, primarily reporting what has been observed for visual perceptual learning, as well as describe the emerging evidence for multisensory perceptual learning.

### *Concepts and Models of Perceptual Learning*

Although many early accounts of perceptual suggested that perceptual learning occurred at low-level sensory areas (Karni & Sagi, 1991; Yotsumoto, Watanabe, & Sasaki, 2008), a number of studies have recently reported that perceptual learning occurs also at higher-order sensory processing regions (C.-T. Law & Gold, 2008; C.-T. Law & Gold, 2009; Sathian, Deshpande, & Stilla, 2013). Increasing evidence suggests that for some forms, of perceptual learning, sensory representations do not change but rather improved performance results from changes in how sensory evidence is weighed to form the decision (C. T. Law & Gold, 2010). One earlier study, in particular, sought to identify the contribution of the middle temporal area (MT), responsive to moving stimuli, and the lateral intraparietal area (LIP), which is strongly predictive of a monkey's choice using electrophysiology with monkeys performing a one-interval, two-alternative direction-discrimination task. Following perceptual training, improvements in performance on the direction-discrimination task corresponded to changes in

the LIP but not the MT. Thus, these results indicated that the perceptual improvements observed for the monkeys on the direction-discrimination task correspond to increasing selectivity of the readout of sensitive MT neurons by a decision process, represented in the LIP (C.-T. Law & Gold, 2008). Computational modeling using a reinforcement-learning rule suggests that changes in LIP during training are due to an increasingly selective readout of the most informative sensory MT neurons (as well as likely other sensory areas), suggestive of changes in the connectivity between sensory and decision representations (C.-T. Law & Gold, 2009). The interaction of low-level and higher-order perceptual processing is likely not unique to visual perceptual learning. Evidence also suggests that top-down and bottom-up signals may interact during auditory perceptual learning in rats (Polley, Steinberg, & Merzenich, 2006).

For visual perceptual learning, a number of perceptual reweighting models have been utilized to account for changes in the readout of a decision unit as well as enhancement of the sensory information (Barbara Anne Doshier & Lu, 1999; B. A. Doshier & Lu, 1998; Petrov, Doshier, & Lu, 2005). Evidence in support of perceptual reweighting models suggests visual perceptual learning occurs as a result of reweighting between the sensory representation and the decision unit as perceptual learning is reported to improve the filtering of external noise and enhance the representation of the stimulus (Barbara Anne Doshier & Lu, 1999; B. A. Doshier & Lu, 1998). The perceptual weighting models have also been observed to account for the influence of feedback on perceptual learning, suggesting that these models can account for sensory, decisional, and feedback factors in perceptual learning (Liu, Doshier, & Lu, 2014).

Another model, the dual plasticity model, has been proposed to account for various types of visual perceptual learning (Watanabe & Sasaki, 2015). Visual perceptual learning can occur for features that are experienced but not specific to task which has been termed task-irrelevant visual perceptual learning (Watanabe, Náñez, & Sasaki, 2001). Subsequently, for visual perceptual learning, improvements in perception have been classified as arising from task-relevant learning (attention focused on the stimuli to be learned) and task-irrelevant

learning (improvements outside the focus of attention). Both task-relevant and task-irrelevant learning are thought to develop from an interaction of diffuse but temporally coincident reinforcement signals and the signals produced by completing a task (A. Seitz & Watanabe, 2005). Task-irrelevant visual perceptual learning has been suggested to occur at low-level sensory areas (Watanabe et al., 2002). A dual plasticity model has subsequently been proposed to suggest that task-relevant learning may elicit changes in task-based plasticity (related to changes in processing a task) and feature-based plasticity (changes in the representation of features) while task-irrelevant learning elicits only changes in feature-based plasticity. Feature-based plasticity is hypothesized to occur at the earlier stages of sensory processing whereas task-based plasticity is hypothesized to occur at the later stages of perceptual processing. Visual perceptual learning that subsequently occurs as a result of interactions between top-down signals and bottom-up sensory signals that is similar to the reweighting models. An advantage of the dual plasticity model is that it allows for feature specific learning at lower sensory levels of processing (Watanabe & Sasaki, 2015).

Another important concept in perceptual learning is the reverse hierarchy theory of visual perceptual learning. The reverse hierarchy theory of visual perceptual learning suggests that the relative difficulty of a task determines at what level in the cortical hierarchy perceptual learning will occur and subsequently how generalizable the improvements in perception will be following perceptual learning. Thus, easier tasks will engage higher-order cortical areas whereas more difficult tasks will elicit perceptual learning in sensory cortex (Ahissar & Hochstein, 2004).

Although many of the discussed models and concepts are derived from studies of unisensory, visual perceptual learning, it is likely that many of these general concepts may also describe multisensory perceptual learning.

### *Feedback and Perceptual Learning*

Although perceptual learning is often elicited by paired feedback with sensory stimuli, practice in the absence of external feedback may be sufficient to elicit perceptual learning. Numerous accounts of visual perceptual learning have been described to occur with and without feedback (Manfred Fahle & Edelman, 1993; M Fahle, Edelman, & Poggio, 1995). Interestingly, perceptual learning in the absence of a feedback signal may be dependent on task difficulty as task difficulty may be either sufficient to elicit perceptual learning or prevent perceptual learning dependent on the presence of feedback. For example, visual perceptual learning may occur with or without feedback when a task is easy but require feedback for visual perceptual learning to occur when a task is difficult (Liu, Lu, & Doshier, 2010). Tasks that are relatively easy have been hypothesized to elicit perceptual learning in the absence of external feedback by evoking an internal reinforcement signal. In some cases, however, easy exemplars (which may elicit reinforcement signals), may not always be sufficient to elicit visual perceptual learning in the absence of a feedback signal (Aaron R. Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006). Interestingly, visual perceptual learning is also observed to occur in the absence of feedback when easy trials and difficult trials (e.g. trials for which an individual responded accurately on 65% or 85% of trials) are mixed while no perceptual learning is observed when only easy trials are presented in the absence off feedback. Furthermore, the magnitude of improvement following mixed difficulty training (i.e. easy and difficult trials) in the absence of feedback is equivalent to perceptual training with difficult trials in the absence of feedback as well as perceptual trailing with easy, difficult, or a mixture of easy and difficult when trial-by-trial feedback is present, suggestive of an internal response generated by the easy trials that allows for perceptual learning (Liu, Lu, & Doshier, 2012).

At some level, perceptual training with or without feedback may be similar as the capacity for each of these training paradigms to elicit perceptual learning is likely dependent on whether the training paradigm can engage reward circuitry. Visual perceptual learning without

feedback may function similar to perceptual training with feedback as confidence in perceptual judgments appears activate the same brain areas that are typically engaged by external feedback (Guggenmos, Wilbertz, Hebart, & Sterzer, 2016). Similarly, it has been suggested that task-irrelevant visual perceptual learning develops due to diffuse reward signals being present while completing a task and also viewing the task-irrelevant visual feature (Aaron R Seitz & Watanabe, 2003). Numerous studies support that the pairing of a sensory with reinforcement is sufficient to elicit perceptual learning. In the absence of a task, the pairing of a reward with a visual feature on a passive visual perceptual learning task (i.e. the participants were not aware of the stimulus-reward contingencies) facilitates perceptual learning for the feature paired with reward while no perceptual learning is observed for a feature that was not paired with a reward. Awareness of the task does not seem to affect this type of perceptual learning as perceptual learning was also observed for presentations of the visual feature that were monocular and imperceptible and paired with reward (and specific to the trained eye) (Aaron R Seitz, Kim, & Watanabe, 2009). Furthermore, visual perceptual learning appears to develop similar to learning following operant conditioning. If a visual stimulus is paired with a positive reward contingency (e.g. 80% probability of reward), perceptual learning is observed for that particular stimulus in comparison to stimuli paired with a 50% probability of reward or 20% (negative reward contingency) probability of reward (D. Kim, Seitz, & Watanabe, 2015). Interestingly, feedback may also inhibit some forms of perceptual learning. Visual perceptual learning that is observed for a stimulus feature paired with feedback may actually inhibit perceptual learning for another feature exposed in the absence of feedback (but still task-relevant) at the time of perceptual training (whereas perceptual learning might normally occur for if both features were presented in the absence of feedback) (Dobres & Watanabe, 2012).

A number of studies have examined whether false or manipulated feedback affects perceptual learning. For perceptual learning to occur, feedback must be generally correlated to responses. For example, if feedback is not correlated to participant responses (i.e. if the

participant makes a correct response, but the probability of receiving feedback indicating a correct response is 0.5), visual perceptual learning does not occur (Michael H Herzog & Fahle, 1997). Similarly, uncorrelated feedback appears to also prevent multisensory perceptual learning (see Chapter 2). Manipulated false feedback (in which incorrect feedback is given on 100% of trials for objectively correct responses) may also elicit perceptual learning, however, the early changes in behavior may be the result of changing criterion prior to changes in sensitivity (Aberg & Herzog, 2012). It is reported that following perceptual training with false feedback, performance rapidly changes when true (objectively correct) feedback is presented instead, suggesting that false (manipulated) external feedback produces a change a change in criterion (M. H. Herzog & Fahle, 1999). Such changes in criterion during visual perceptual learning has been suggested as a strategy for minimizing error feedback quickly and avoiding incorrect perceptual learning as is observed for perceptual learning with reverse feedback (Michael H. Herzog, Ewald, Hermens, & Fahle, 2006). Interestingly, changes in criterion and sensitivity might co-occur during a perceptual training tasks, although changes in sensitivity are likely to be present in a subsequent training session on a second day of perceptual training while changes in criterion will have likely diminished. Additionally, changes in criterion may be dependent on the type of feedback as for visual perceptual learning, trial-by-trial feedback and block feedback are both capable of altering sensitivity but trial-by-trial feedback may also change criterion while changes in criterion are typically not observed with block feedback (Aberg & Herzog, 2012).

Fake feedback may also facilitate visual perceptual learning as fake feedback that indicates a larger improvement in performance than expected facilitates perceptual learning compared to genuine feedback. Conversely, fake feedback that indicates worse than expected performance produces similar visual perceptual learning to visual perceptual learning that occurs with genuine feedback, suggesting that participants may be likely to discount the more negative fake feedback. Collectively, these results suggest that some forms of perceptual

learning may be influenced by more cognitive processes (Shibata, Yamagishi, Ishii, & Kawato, 2009).

Perceptual training with feedback may also lead to more durable perceptual learning. A study by *Dobres & Watanabe (2012)* demonstrated that extensive training with feedback maintains performance on a 2-IFC motion detection task whereas perceptual learning that occurs in the absence of feedback or with less extensive perceptual training with feedback deteriorates with increasing exposure to non-trained motion direction stimuli in a subsequent post-training motion detection test. These findings are suggestive that in the absence of extensive training with feedback, perceptual learning is subject to deterioration with additional exposure to differing motion stimuli, suggesting that a feedback signal is necessary during training for consolidation of perceptual learning that is durable (Dobres & Watanabe, 2012).

Although the contribution of feedback to perceptual learning has been extensively studied, the majority of these studies have focused primarily on the role of feedback for visual perceptual learning. Feedback may drive multisensory perceptual learning either differently or similarly to visual perceptual learning. For example, feedback may be necessary to elicit multisensory perceptual learning. It is possible that the valence of feedback (i.e. positive or negative feedback) may also influence multisensory perceptual learning. Neural evidence suggests that positive (i.e. response was correct) and negative (i.e. response was incorrect) feedback utilize both common as well as separate neural networks (Bischoff-Grethe, Hazeltine, Bergren, Ivry, & Grafton, 2009). Thus, many of the studies described in the following chapters will address how feedback influences multisensory perceptual learning.

### *Generalization of Perceptual Learning*

Perceptual learning, particularly visual perceptual learning, is often reported to be very specific to stimulus features (Lu, Hua, Huang, Zhou, & Doshier, 2011). Initial studies of perceptual learning often reported that perceptual learning at one retinal location (i.e. portion of

the visual field) would not transfer to another retinal location (M Fahle et al., 1995; Karni & Sagi, 1991). Although early studies suggested that visual perceptual learning was specific to retinal location, it was later observed that performing a different, secondary task at the location would facilitate transfer of perceptual learning to the location of the secondary task (perceptual learning for a feature at one location transfers to a second location where perceptual learning occurred for a different feature). This perceptual training paradigm, known as double training, elicits transfer of perceptual learning if the secondary task is performed prior to the primary task or if the primary and secondary task blocks of training trials are interleaved (Xiao et al., 2008). Double-training procedures require the secondary task to be task-relevant (i.e. passive presentation of another stimulus does not enable transfer of perceptual learning) for visual perceptual learning to occur (Mastropasqua, Galliussi, Pascucci, & Turatto, 2015). In some sense, as perceptual learning for the secondary task also transfers to the location of the primary task, neither task is really primary or secondary. Interestingly, more extensive double-training is required for transfer to occur as fewer tasks containing the secondary task training trials does not facilitate transfer of the primary task to a different location (Mastropasqua et al., 2015). The generalization of visual perceptual learning, however, may still not generalize to more general aspects of visual perception. For example, training-plus-exposure (which essentially is similar to double training) is not capable of enabling the transfer of different visual feature (i.e. training with contrast discrimination does not enhance orientation discrimination), suggestive that the specificity of visual perceptual learning occurs because a specific set of rules is learned for a specific task (Cong, Wang, Yu, & Zhang, 2016).

Visual perceptual learning may also generalize if the difficulty of a task is relatively easy (while in comparison difficult tasks often elicit specific visual perceptual learning). The generalization of visual perceptual learning that occurs with easy tasks is hypothesized to occur as a result of visual perceptual learning occurring in higher visual regions while specific visual perceptual learning elicited by difficult tasks occurs at lower visual regions (Ahissar & Hochstein,

1997). More recent evidence from visual perceptual learning suggests that the difficulty, or rather precision, of the transfer stimuli determine whether transfer will occur rather than overall task difficulty, particularly the difficulty of the training task. Thus, whether the training task is initially for a high precision or low precision stimulus (i.e. relative difficulty is equated, but the differences in the stimulus features are either more or less precise), transfer will only occur to a low precision stimulus rather than a high precision stimulus (Jeter, Doshier, Petrov, & Lu, 2009). Additionally, fewer training trials appear to enable better transfer of visual perceptual learning to other locations and stimulus orientations, also suggesting that the capacity for the transfer perceptual learning decreases with increasing specificity of perceptual learning (Jeter, Doshier, Liu, & Lu, 2010).

One theory, the integrated reweighting theory of perceptual learning and transfer, posits that visual perceptual learning occurs simultaneously at multiple levels of the cortical hierarchy such that a reweighting of higher-level regions that represent location-independent stimulus features enables retinal location transfer while stimulus feature transfer is accomplished by reweighting of location-specific and location-independent representations (Barbara Anne Doshier, Jeter, Liu, & Lu, 2013). Thus, generalization observed following double-training or with perceptual training with easy tasks may better engage regions of perceptual decision making and thus enable generalization of visual perceptual learning.

Perceptual learning may exhibit greater capacity for generalization of improved perception of amodal features across sensory modalities. For example, it was recently reported that perceptual training on either a unisensory spatial or temporal task with visual or auditory stimuli elicits significant perceptual learning for the specifically trained task, however, the effects of perceptual training only generalize on the spatial task from visual spatial training to improved auditory spatial function and on the temporal task from auditory temporal training to improved visual temporal function. Interestingly, the generalization reported in these findings suggests that transfer across sensory modalities occurs from the sensory modality that excels on a

specific task (in this case, vision being the dominant sensory modality for spatial tasks and audition being the dominant sensory modality for temporal tasks) to a different sensory modality (McGovern, Astle, Clavin, & Newell, 2016). Within modality training may not always result in perceptual learning. Occasionally, training in different sensory modality (that is more proficient at a particular task) may be necessary to elicit perceptual learning in a different sensory modality. For example, visual rhythm training fails to improve visual rhythm discrimination while auditory and audiovisual rhythm training improve visual rhythm discrimination<sup>5</sup> (Barakat, Seitz, & Shams, 2015). Visual signals may also facilitate auditory perceptual learning as auditory perceptual learning on a monaural auditory localization task is enhanced by audiovisual training (practice) in comparison to auditory only training (practice) or auditory training with feedback (where feedback was given after practice with some auditory only trials) (Strelnikov, Rosito, & Barone, 2011).

Some forms of perceptual learning may also only generalize after audiovisual training as a study that employed training (with no apparent overt feedback) on a visual, auditory, or audiovisual TOJ tasks observed that there was improvement following training on the trained modality but only the audiovisual training generalized to improvements for the visual-only TOJ task. Furthermore, it was observed that both visual and audiovisual TOJ training produced generalized improvements when a feature of the visual stimulus was changed or if the location of the visual stimulus changed. Improvements on the auditory TOJ task did not result in generalized improvements for auditory tones. While training improved temporal acuity for stimulus onset discrimination in the TOJ task (participants were only responded to onsets for TOJ tasks during training), a generalized improvement in temporal acuity for stimulus offset discrimination was not observed (Alais & Cass, 2010).

---

<sup>5</sup> In this study, although auditory and audiovisual rhythm training each improved visual rhythm discrimination, there was no additional benefit of audiovisual training.

Perceptual learning in a sensory modality following multisensory perceptual training that results in greater learning compared to unisensory perceptual training may be due to the capacity for multisensory perceptual training to activate higher order multisensory regions. Recent neural evidence of improved visual motion coherence discrimination following audiovisual training with congruent acoustic textures (compared to visual training or visual training with auditory noise) implicates the selective recruitment of multisensory regions following training that facilitate visual processing (Zilber, Ciuciu, Gramfort, Azizi, & Van Wassenhove, 2014). These results support the hypothesis that supramodal processing in higher-order cortices facilitates the generalization of perceptual learning as well as influences the tuning in sensory cortex (Zilber et al., 2014), consistent with the reverse hierarchy of theory of visual perceptual learning (Ahissar & Hochstein, 2004).

#### *Multisensory Perceptual Learning and Temporal Acuity*

Recently, there has been an interest in understanding perceptual learning in a multisensory context. Numerous efforts have addressed the facilitation of unisensory perceptual learning by multisensory perceptual training as well as described perceptual learning for distinctly multisensory stimuli (i.e. features of multisensory stimuli). Although multisensory perceptual training may enhance generalization of unisensory perceptual learning, multisensory perceptual training may also result in better performance, particularly for multisensory stimuli. For example, audiovisual perceptual training (in which the auditory stimulus could be perceived as directional auditory motion) results in better performance on a motion-detection task within and across training sessions compared to visual perceptual training (Aaron R Seitz et al., 2006). Further evidence suggests that the facilitation of unisensory perceptual learning by multisensory training appears to possibly result from an involvement of higher order multisensory regions during perceptual learning. Interestingly, the auditory stimulus only enhances perceptual learning if it is presented in congruency with the visual stimulus (i.e. the auditory and visual

motion cues are in the same direction). Thus, when individuals are asked in a 2-IFC task if a directional stimulus was present, congruent audiovisual cues enhance perceptual learning while incongruent audiovisual cues only elicit marginal perceptual learning despite the fact that both congruent and incongruent stimuli provide evidence of directionality (R. S. Kim, Seitz, & Shams, 2008). Such findings suggest that multisensory interactions likely facilitate the improved performance after multisensory perceptual training.

Not many studies of multisensory perceptual learning have examined perceptual learning for specific multisensory features, although the few studies that have been conducted have examined how multisensory perceptual learning may improve temporal acuity, particularly for audiovisual stimuli. Enhanced temporal acuity was first observed by *Powers et al. (2009)* who reported that the TBW for audiovisual “flash-beep” stimuli was significantly narrower following five days of multisensory perceptual training (Powers et al., 2009). In this study, both 2-AFC (yes/no) and 2-IFC SJ tasks were utilized for perceptual training tasks as well as to estimate the TBW prior to and following each perceptual training session. As the TBW was observed to narrow following training on both tasks, this suggests that changes in temporal acuity occurred due to perceptual changes rather than cognitive biases. Furthermore, the changes in the TBW reported by this study are likely to result from perceptual learning as the study additionally reported changes in temporal acuity were observed to remain one week after the last perceptual training session. In comparison, no changes in the TBW were observed for individuals performing the 2-AFC training paradigm in the absence of feedback while the TBW subjects passively exposed to asynchronous stimuli may actually become wider. Collectively, these findings suggest that multisensory perceptual training is capable of enhancing temporal acuity for audiovisual stimuli. The findings of *Powers et al. (2009)* have subsequently been replicated by our own lab (Powers, Hevey, & Wallace, 2012) as well as others using different training tasks (like a TOJ task) (Setti et al., 2014).

Multisensory perceptual training to enhance temporal acuity may also generalize to affect unisensory temporal acuity as well as measures of multisensory integration. Two studies have explored whether changes in temporal acuity following perceptual training generalize to affect the perception of the SIFI. One study observed that while performance on the SIFI following perceptual training did not change, visual temporal acuity was enhanced following perceptual training as evidenced by increased correct responses of the number of flashes for trials for which two flashes were presented (Powers III, Hillock-Dunn, & Wallace, 2016). In another study, it was observed that multisensory perceptual training that improved temporal acuity (i.e. for only the individuals for who perceptual training improved temporal acuity) also reduced susceptibility to the SIFI (Setti et al., 2014). The reduction in susceptibility to the SIFI suggests that the magnitude of multisensory integration may be affected following multisensory perceptual training as smaller TBWs are reported to correlate with reduced perception of the SIFI (Ryan A. Stevenson, Zemtsov, et al., 2012). As these studies have primarily used “flash-beep” like stimuli, it remains unknown whether improvement in temporal acuity may generalize to other multisensory tasks.

A recent study also sought to examine the effects of multisensory perceptual learning on spatiotemporal acuity following perceptual training to enhance temporal acuity. In this study, it was reported that perceptual training to enhance temporal acuity elicited narrowing of the TBW similar to previous studies. The increase in temporal acuity following perceptual training was not, however, observed to generalize to affect spatial integration. Using a Bayesian causal inference model, this study further revealed that the effects of perceptual training could be accounted for by an increase in the precision of temporal acuity for audiovisual stimuli as well as a decrease in the prior belief that the stimuli derive from a common source. Interestingly, the change in temporal acuity is hypothesized to be distinct from the change in expectations that the auditory and visual stimuli derive from a common source (McGovern, Roudaia, et al., 2016). Such findings indicate that future studies will be necessary to determine the generalizability of

multisensory perceptual learning as well as if perceptual representations and prior beliefs are affected by multisensory perceptual training.

### *Neural Evidence of Perceptual Learning*

Although less extensive in comparison to behavioral studies, a number of studies have sought to address the neural basis of perceptual learning. For visual perceptual learning, some evidence has supported suggesting that visual perceptual learning occurs at the early stages of processing in visual cortex (Schwartz, Maquet, & Frith, 2002). Others have observed changes associated with visual perceptual learning in higher order regions (Kahnt, Grueschow, Speck, & Haynes, 2011; Shibata, Sasaki, Kawato, & Watanabe, 2016). One study in particular observed that for visual perceptual learning, stimulus features were encoded at lower level sensory areas such as early visual cortex as well as higher order cortical regions such as the lateral intraparietal area (LIP). Yet only learning-related changes in the task were found in the anterior cingulate cortex (ACC) that encoded learning-related changes in the decision variable in comparison to the stimulus feature. Additionally reward prediction-error signaling in reward-related and higher decision-making areas such as the left ventral striatum, bilateral anterior insular cortex extending into the lateral OFC, the dorsolateral PFC, and the dorsomedial prefrontal cortex including the ACC (Kahnt et al., 2011). Thus, it is likely that perceptual learning elicits neural changes in multiple neural regions.

Evidence in support of the dual plasticity model of perceptual learning suggests that feature-based and task-based plasticity may be present following visual perceptual learning, which may account for different changes in different neural regions. Following training on a motion detection that elicited visual perceptual for the trained motion direction, improved pattern-classification analysis (reflective of how precisely the patterns of the fMRI responses in each region reflect a particular motion stimulus) was observed for V3A when motion stimuli (for the trained motion direction) were presented either passively or as part of a motion detection

task while improved pattern-classification analysis was only observed for V1 and the intraparietal sulcus (IPS) when motion stimuli were presented as part of a motion detection task. Thus, the changes in V3A following perceptual training are likely the result of feature-based plasticity whereas the changes in V1 and the IPS are the result of task-based plasticity (Shibata et al., 2016). It seems likely that changes in sensory and decisional regions both contribute to visual perceptual learning.

One of the first descriptions of the neural basis of multisensory perceptual learning identified changes in both low and high level sensory regions (Powers et al., 2012). Using a similar perceptual training paradigm as Powers et al. (2009), multisensory perceptual learning was reported to elicit a decrease in activity in the pSTS as well as visual and auditory cortices, suggestive of enhanced neural efficiency. Additionally, effective connectivity following multisensory perceptual training is observed to change from a feedforward network from auditory and visual cortex to pSTS to a more distributed network involving feedback to visual and auditory cortex (Powers et al., 2012). Thus, multisensory perceptual learning likely involves changes in multisensory regions, such as the pSTS, and may also affect other sensory regions. Multisensory training may also enhance unisensory perception through a supramodal network that includes the pSTS. A recent study observed that audiovisual training with coherent audiovisual textures increased activation of the pSTS, mSTS, and auditory cortex (presumably primary auditory cortex) while audiovisual training with auditory noise decreased activity in these same areas. The pattern of activity observed in this study were interpreted to suggest that following audiovisual perceptual training with coherent auditory textures, the pSTS enabled the continued flow of information from mSTS to hMT+ that enhanced visual processing, indicative of a supramodal enhancement of visual processing (Zilber et al., 2014).

It is also important to note that the changes in neural activity observed with perceptual learning may change as learning progresses. Evidence suggests that there are distinct temporal phases in the changes in activity in early sensory cortex (i.e. V1) that accompany visual

perceptual learning. In the early period of visual perceptual learning, V1 activity increases yet as the level of performance was maintained, activity in V1 decreased to pre-training levels. The observed decrease in V1 activity is suggestive of possible increasing neural efficiency such that the areas of V1 that are activated during perceptual learning become more specific (Yotsumoto et al., 2008). Thus, when considering the changes in neural activity observed by some studies that have examined neural changes following extensive training (Powers et al., 2012), a full account of the neural changes associated with perceptual learning may be incomplete.

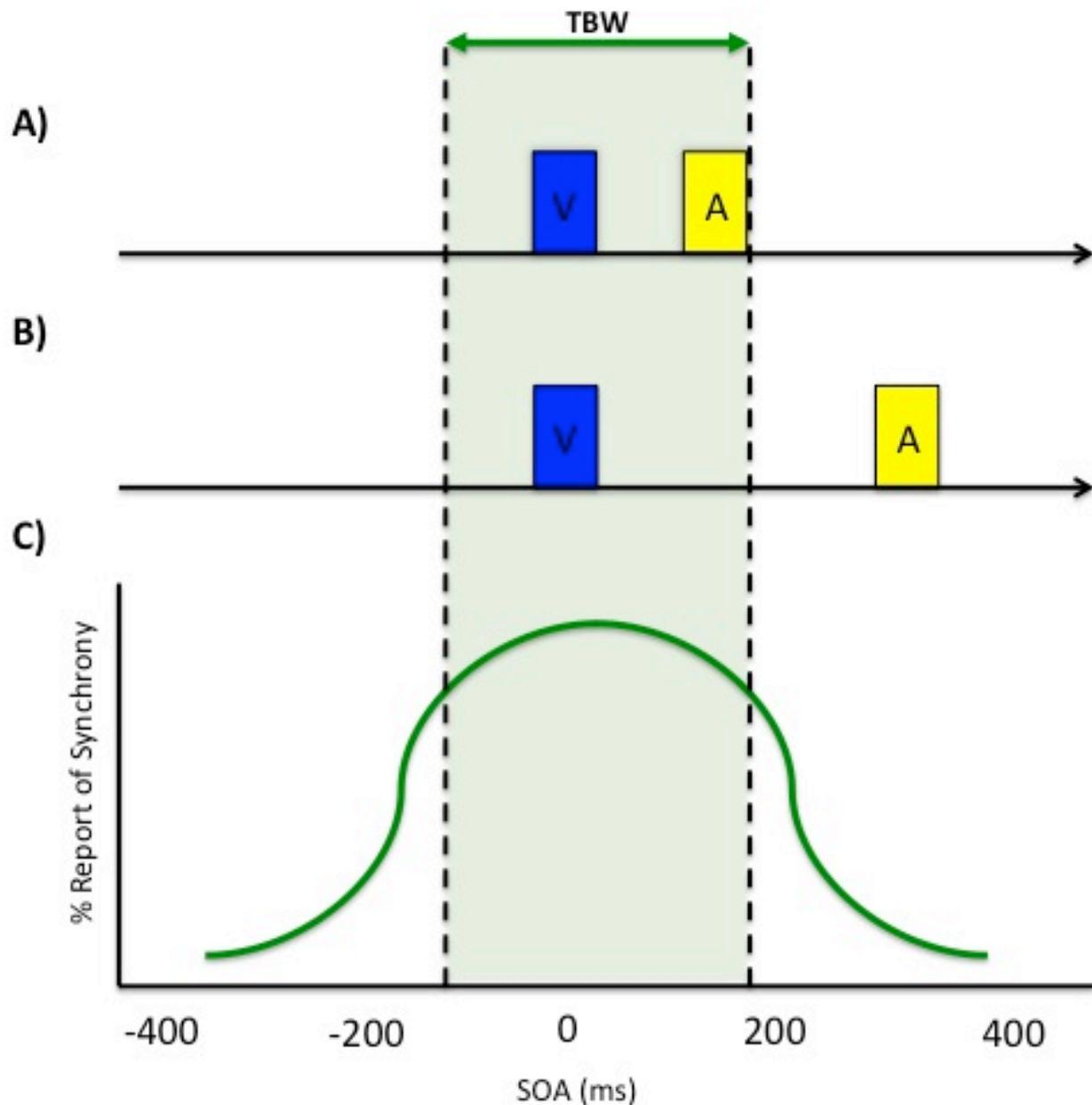
### **Introduction to the Subsequent Chapters**

The following sections will describe studies designed to further our understanding of multisensory plasticity, particularly the plasticity of the perceptual representations of audiovisual temporal relationships. The work described in Chapters 2 and 3 examines the contingencies with which multisensory perceptual learning develops following perceptual training. In particular, Chapter 2 explores the relationship of multisensory perceptual learning and feedback during perceptual training while Chapter 3 explores the effect of task difficulty on multisensory perceptual learning. In Chapter 4, the influence of feedback on the time course of cumulative and rapid temporal recalibration is explored. Chapter 4 provides evidence that top-down and bottom-up interactions likely facilitate changes in the perceptual representations of multisensory temporal relationships. The findings in Chapter 5 demonstrate that perceptual training is also capable of enhancing temporal acuity for audiovisual speech stimuli, however, changes in temporal acuity that are elicited by perceptual training do not appear to generalize across different levels of stimulus complexity. Finally, in Chapter 6 we report that perceptual training to enhance temporal acuity for audiovisual speech results in widespread changes neural activity following perceptual training and may elicit changes in perceptual decision making regions.

Collectively, these findings further our understanding of multisensory perceptual learning as well as multisensory temporal acuity. Importantly, temporal acuity has been observed to be

impaired in a number of neurodevelopmental disorders, including autism spectrum disorder (ASD) (Wallace & Stevenson, 2014). In particular, it has been reported that individuals with ASD may have impaired temporal acuity for audiovisual speech stimuli (Ryan A Stevenson, Siemann, et al., 2014). Subsequently, perceptual training to enhance temporal acuity has been suggested as a possible strategy to enhance temporal acuity for individuals for whom temporal acuity is impaired (Wallace & Stevenson, 2014). The findings in Chapters 5 and 6 that demonstrate that perceptual training is capable of enhancing temporal acuity for audiovisual speech for neurotypical adults suggest that perceptual training may also be able to enhance impaired temporal acuity for audiovisual speech. Although, future investigations will be necessary to determine if perceptual training can enhance temporal acuity for those for whom temporal acuity is impaired, the findings of the following chapters provide evidence that perceptual training is at least capable of enhancing temporal acuity in a typically-developed adult population.

The discussion will review some of the more prominent findings of the previously described studies. A number of themes pertinent to the investigations and findings of the previous chapters will be discussed in more detail. The contribution of the interaction of bottom-up and top-down to multisensory perceptual learning will be explored. As experience and prior expectations are observed to significantly influence multisensory processing (Gau & Noppeney, 2016), the changes in multisensory representations that occur with learning and perceptual training will be considered. Additionally, considerations for future investigations and implications of the findings described by the previous chapters for neurodevelopmental disorders such as ASD will also be discussed.



**Figure 1-1. The Temporal Principle of Multisensory Integration** A) Sensory cues from separate sensory modalities (i.e auditory and visual stimuli) that are proximate in time are likely to be perceptually bound. The epoch of time for which auditory and visual stimuli may be asynchronous but still perceptually bound is described as the temporal binding window (TBW). B) When the auditory and visual stimuli become increasingly disparate in time the sensory stimuli are less likely to be perceptually bound and rather are likely to be perceived as separate sensory events. C) Participants presented with auditory and visual stimuli at various stimulus onset asynchronies (SOAs) for a simultaneity judgment (SJ) task will often report the audiovisual stimuli to be synchronous over an epoch of SOAs. The SOAs for which participants reliably perceive the stimuli to be synchronous can be used to estimate the TBW. V = visual (blue), A = auditory (yellow).

## References

- Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, 12(3). doi:10.1167/12.3.3
- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631), 401-406.
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences*, 8(10), 457-464.
- Alais, D., & Cass, J. (2010). Multisensory perceptual learning of temporal order: audiovisual learning transfers to vision but not audition. *Plos One*, 5(6), e11283.
- Alais, D., Orchard-Mills, E., & Van der Burg, E. (2015). Auditory frequency perception adapts rapidly to the immediate past. *Attention, Perception, & Psychophysics*, 77(3), 896-906.
- Alcalá-Quintana, R., & García-Pérez, M. A. (2013). Fitting model-based psychometric functions to simultaneity and temporal-order judgment data: MATLAB and R routines. *Behavior research methods*, 45(4), 972-998.
- Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision research*, 48(25), 2537-2544.
- Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, 92(2), 339-355.
- Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, 6(3), 6-6.
- Bachmann, T., Pöder, E., & Luiga, I. (2004). Illusory reversal of temporal order: The bias to report a dimmer stimulus as the first. *Vision research*, 44(3), 241-246.
- Balk, M. H., Ojanen, V., Pekkola, J., Autti, T., Sams, M., & Jääskeläinen, I. P. (2010). Synchrony of audio–visual speech stimuli modulates left superior temporal sulcus. *Neuroreport*, 21(12), 822-826.

- Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, 25(2), R60-R61.  
doi:10.1016/j.cub.2014.12.011
- Barrós-Loscertales, A., Ventura-Campos, N., Visser, M., Alsius, A., Pallier, C., Rivera, C. Á., & Soto-Faraco, S. (2013). Neural correlates of audiovisual speech processing in a second language. *Brain and language*, 126(3), 253-262.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature neuroscience*, 7(11), 1190-1192.
- Beauchamp, M. S., Yasar, N. E., Frye, R. E., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *NeuroImage*, 41(3), 1011-1020.
- Benjamins, J. S., van der Smagt, M. J., & Verstraten, F. A. (2008). Matching auditory and visual signals: Is sensory modality just another feature? *Perception*, 37(6), 848-858.
- Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R. B., & Grafton, S. T. (2009). The influence of feedback valence in associative learning. *NeuroImage*, 44(1), 243-251.
- Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research*, 198(2-3), 233-244.
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, 14(2), 427-438.
- Carriere, B. N., Royal, D. W., Perrault, T. J., Morrison, S. P., Vaughan, J. W., Stein, B. E., & Wallace, M. T. (2007). Visual deprivation alters the development of cortical multisensory integration. *Journal of neurophysiology*, 98(5), 2858-2867.
- Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience*, 16(6), 1000-1009.

- Cong, L.-J., Wang, R.-J., Yu, C., & Zhang, J.-Y. (2016). Perceptual learning of basic visual features remains task specific with Training-Plus-Exposure (TPE) training. *Journal of Vision, 16*(3), 13-13.
- Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The journal of the acoustical society of america, 119*(6), 4065-4073.
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics, 66*(8), 1388-1404.
- Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: Relating reaction times and judgments of temporal order. *Psychological review, 122*(2), 232.
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.
- Dobres, J., & Watanabe, T. (2012). Response feedback triggers long-term consolidation of perceptual learning independently of performance gains. *Journal of Vision, 12*(8), 9.
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Research, 1242*, 136-150.
- Doshier, B. A., Jeter, P., Liu, J., & Lu, Z.-L. (2013). An integrated reweighting theory of perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 110*(33), 13678-13683. doi:10.1073/pnas.1312552110
- Doshier, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision research, 39*(19), 3197-3221.
- Doshier, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America, 95*(23), 13988-13993. doi:10.1073/pnas.95.23.13988

- Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Frontiers in psychology*, 6. doi:10.3389/fpsyg.2015.00736
- Erickson, L. C., Zielinski, B. A., Zielinski, J. E., Liu, G., Turkeltaub, P. E., Leaver, A. M., & Rauschecker, J. P. (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect. *Frontiers in psychology*, 5, 534.
- Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Current opinion in neurobiology*, 15(2), 154-160.
- Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision research*, 33(3), 397-412.
- Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision research*, 35(21), 3003-3013.
- Fister, J. K., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature neuroscience*, 7(7), 773-778.
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & psychophysics*, 68(7), 1191-1203.
- García-Pérez, M. A., & Alcalá-Quintana, R. (2011). Interval bias in 2AFC detection tasks: sorting out the artifacts. *Attention, Perception, & Psychophysics*, 73(7), 2332-2352.
- García-Pérez, M. A., & Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: a quantitative model. *Psychonomic Bulletin & Review*, 19(5), 820-846.
- Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, 124, 876-886.
- Ghose, G. M. (2004). Learning in mammalian sensory cortex. *Current opinion in neurobiology*, 14(4), 513-518.

- Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, 49(1), 585-612.
- Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *Elife*, 5, e13388.
- Harvey, C., Van der Burg, E., & Alais, D. (2014). Rapid temporal recalibration occurs crossmodally without stimulus specificity but is absent unimodally. *Brain Research*, 1585, 120-130. doi:10.1016/j.brainres.2014.08.028
- Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of experimental psychology*, 63(3), 289.
- Herzog, M. H., Ewald, K. R. F., Hermens, F., & Fahle, M. (2006). Reverse feedback induces position and orientation specific changes. *Vision research*, 46(22), 3761-3770. doi:10.1016/j.visres.2006.04.024
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision research*, 37(15), 2133-2141.
- Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision research*, 39(25), 4232-4243. doi:10.1016/s0042-6989(99)00138-8
- Hirsh, I. J., & Sherrick Jr, C. E. (1961). Perceived order in different sense modalities. *Journal of experimental psychology*, 62(5), 423.
- Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *Plos One*, 4(7), e6450.
- Jeter, P. E., Doshier, B. A., Liu, S.-H., & Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision research*, 50(19), 1928-1940.
- Jeter, P. E., Doshier, B. A., Petrov, A., & Lu, Z.-L. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision*, 9(3), 1-1.

- Kahnt, T., Grueschow, M., Speck, O., & Haynes, J.-D. (2011). Perceptual Learning and Decision-Making in Human Medial Frontal Cortex. *Neuron*, *70*(3), 549-559.  
doi:10.1016/j.neuron.2011.02.054
- Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966-4970.
- Karni, A., & Sagi, D. (1993). The time course of learning a visual skill.
- Kayser, C., & Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biol*, *13*(2), e1002075.
- Keetels, M., & Vroomen, J. (2007). No effect of auditory–visual spatial disparity on temporal recalibration. *Experimental Brain Research*, *182*(4), 559-565.
- Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile–visual asynchronous stimuli. *Neuroscience letters*, *430*(2), 130-134.
- Kim, D., Seitz, A. R., & Watanabe, T. (2015). Visual perceptual learning by operant conditioning training follows rules of contingency. *Visual Cognition*(ahead-of-print), 1-14.
- Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *Plos One*, *3*(1), e1532.
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *NeuroImage*, *37*(4), 1445-1456.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405-414.
- Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat Neurosci*, *11*(4), 505-513.  
doi:[http://www.nature.com/neuro/journal/v11/n4/supinfo/nn2070\\_S1.html](http://www.nature.com/neuro/journal/v11/n4/supinfo/nn2070_S1.html)

- Law, C.-T., & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature neuroscience*, *12*(5), 655-663.
- Law, C. T., & Gold, J. I. (2010). Shared mechanisms of perceptual learning and decision making. *Topics in cognitive science*, *2*(2), 226-238.
- Lee, H., & Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *The Journal of Neuroscience*, *31*(31), 11338-11350.
- Lee, H., & Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. *Current Biology*, *24*(8), R309-R310.
- Liu, J., Doshier, B., & Lu, Z.-L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision research*, *99*, 46-56. doi:10.1016/j.visres.2014.01.001
- Liu, J., Lu, Z.-L., & Doshier, B. A. (2010). Augmented Hebbian reweighting: Interactions between feedback and training accuracy in perceptual learning. *Journal of Vision*, *10*(10), 29-29.
- Liu, J., Lu, Z.-L., & Doshier, B. A. (2012). Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision research*, *61*, 15-24.  
doi:10.1016/j.visres.2011.12.002
- Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *Plos One*, *8*(1), e54798.
- Lu, Z.-L., Hua, T., Huang, C.-B., Zhou, Y., & Doshier, B. A. (2011). Visual perceptual learning. *Neurobiology of learning and memory*, *95*(2), 145-151.
- Machulla, T.-K., Di Luca, M., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, *217*(1), 89-97.
- MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British journal of audiology*, *21*(2), 131-141.

- Magnotti, J. F., Ma, W. J., & Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. *Frontiers in psychology*, 4, 798.
- Marchant, J. L., & Driver, J. (2012). Visual and audiovisual effects of isochronous timing on visual perception and brain activity. *Cerebral Cortex*, bhs095.
- Marchant, J. L., Ruff, C. C., & Driver, J. (2012). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target - detection performance for both modalities. *Human brain mapping*, 33(5), 1212-1224.
- Mastropasqua, T., Galliussi, J., Pascucci, D., & Turatto, M. (2015). Location transfer of perceptual learning: Passive stimulation and double training. *Vision research*, 108, 93-102.
- McGovern, D. P., Astle, A. T., Clavin, S. L., & Newell, F. N. (2016). Task-specific transfer of perceptual learning across sensory modalities. *Current Biology*, 26(1), R20-R21.
- McGovern, D. P., Roudaia, E., Newell, F. N., & Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific reports*, 6, 24673.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices.
- Mendonça, C., Santos, J. A., & López-Moliner, J. (2011). The benefit of multisensory integration with biological motion signals. *Experimental Brain Research*, 213(2-3), 185-192.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, 7(10), 3215-3229.
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *science*, 221(4608), 389-391.
- Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350-354.

- Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology*, 56(3), 640-662.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive psychology*, 14(2), 247-279.
- Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience*, 25(25), 5884-5893.  
doi:10.1523/jneurosci.0896-05.2005
- Miller, R. L., Pluta, S. R., Stein, B. E., & Rowland, B. A. (2015). Relative unisensory strength and timing predict their multisensory product. *The Journal of Neuroscience*, 35(13), 5213-5220.
- Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends in neurosciences*.
- Navarra, J., Alsius, A., Velasco, I., Soto-Faraco, S., & Spence, C. (2010). Perception of audiovisual speech synchrony for native and non-native language. *Brain Research*, 1323, 84-93.
- Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences*, 106(23), 9169-9173.
- Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience letters*, 413(1), 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. *Cognitive brain research*, 25(2), 499-507.

- Nidiffer, A. R., Stevenson, R. A., Fister, J. K., Barnett, Z. P., & Wallace, M. T. (2016). Interactions between space and effectiveness in human multisensory performance. *Neuropsychologia*.
- Noel, J.-P., Łukowska, M., Wallace, M., & Serino, A. (2016). Multisensory simultaneity judgment and proximity to the body. *Journal of Vision*, 16(3), 21-21.
- Noel, J.-P., Wallace, M. T., Orchard-Mills, E., Alais, D., & Van der Burg, E. (2015). True and Perceived Synchrony are Preferentially Associated With Particular Sensory Pairings. *Scientific reports*, 5.
- Noesselt, T., Bergmann, D., Münte, T. F., Heinze, H.-J., & Spence, C. (2012). Coding of multisensory temporal patterns in human superior temporal sulcus. *Frontiers in Integrative Neuroscience*, 6, 64.
- Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *The Journal of Neuroscience*, 27(42), 11431-11441.
- Odegaard, B., & Shams, L. (2016). The Brain's Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological science*.
- Olasagasti, I., Bouton, S., & Giraud, A.-L. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex*.
- Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., & Pollick, F. E. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Experimental Brain Research*, 198(2-3), 339-352.
- Petrini, K., Holt, S. P., & Pollick, F. (2010). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *Journal of Vision*, 10(5), 2-2.

- Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, 110(3), 432-439.
- Petrov, A. A., Doshier, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological review*, 112(4), 715.
- Philippi, T. G., van Erp, J. B., & Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment. *Brain Research*, 1242, 116-125.
- Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *The Journal of Neuroscience*, 26(18), 4970-4982.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, 32(18), 6263-6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, 29(39), 12265-12274.
- Powers III, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports*, 6, 23374.
- Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1710), 1314-1322. doi:10.1098/rspb.2010.1737
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17(5), 1147-1153.
- Sathian, K., Deshpande, G., & Stilla, R. (2013). Neural changes with tactile learning reflect decision-level reweighting of perceptual readout. *The Journal of Neuroscience*, 33(12), 5387-5398.

- Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences*, *99*(26), 17137-17142.
- Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in cognitive sciences*, *9*(7), 329-334.
- Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current opinion in neurobiology*, *17*(2), 148-153.
- Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron*, *61*(5), 700-707.
- Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, *16*(14), 1422-1427.
- Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*(10), 1445-1453.
- Seitz, A. R., Nanez, J. E., Sr., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, *6*(9), 966-973. doi:10.1167/6.9.9
- Seitz, A. R., & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature*, *422*(6927), 36-36.
- Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia*, *61*, 259-268.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive brain research*, *14*(1), 147-152.

- Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923-1927.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, *12*(11), 411-417.
- Shibata, K., Sasaki, Y., Kawato, M., & Watanabe, T. (2016). Neuroimaging Evidence for 2 Types of Plasticity in Association with Visual Perceptual Learning. *Cerebral Cortex*, bhw176.
- Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision research*, *49*(21), 2574-2585.
- Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk–MacDonald illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 580.
- Spence, C. (2015). Multisensory flavor perception. *Cell*, *161*(1), 24-35.
- Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., . . . Sathian, K. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*(10), 1713-1720.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255-266.
- Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault Jr, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, *198*(2-3), 113-126.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, *15*(8), 520-535.
- Stein, B. E., & Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Progress in brain research*, *112*, 289-299.

- Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, *49*(4), 3308-3318.
- Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research*, *219*(1), 121-137. doi:10.1007/s00221-012-3072-1
- Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., . . . Wallace, M. T. (2014). Identifying and quantifying multisensory integration: a tutorial review. *Brain topography*, *27*(6), 707-730.
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*, *34*(3), 691-697.
- Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., & James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*, *55*(3), 1339-1345.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249-261. doi:10.1007/s00221-013-3507-3
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. *Journal of Experimental Psychology-Human Perception and Performance*, *38*(6), 1517-1529. doi:10.1037/a0027339
- Strelnikov, K., Rosito, M., & Barone, P. (2011). Effect of audiovisual training on monaural spatial hearing in horizontal plane. *Plos One*, *6*(3), e18344.
- Sumbly, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26*(2), 212-215.

- Tsilionis, E., & Vatakis, A. (2016). Multisensory binding: is the contribution of synchrony and semantic congruency obligatory? *Current Opinion in Behavioral Sciences*.
- Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception & psychophysics*, 42(3), 224-239.
- Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid Recalibration to Audiovisual Asynchrony. *Journal of Neuroscience*, 33(37), 14633-14637. doi:10.1523/jneurosci.1182-13.2013
- Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific reports*, 5.
- Van der Burg, E., Cass, J., Olivers, C. N., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *Plos One*, 5(5), e10664.
- Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B-Biological Sciences*, 282(1804). doi:10.1098/rspb.2014.3083
- Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1053.
- Van der Burg, E., Orchard-Mills, E., & Alais, D. (2015). Rapid temporal recalibration is unique to audiovisual stimuli. *Experimental Brain Research*, 233(1), 53-59.
- Van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & psychophysics*, 70(6), 955-968.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.
- Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the “unity effect” reveals that speech is special. *Journal of Vision*, 8(9), 14-14.

- Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Experimental Brain Research*, *185*(3), 521-529.
- Vatakis, A., & Spence, C. (2006a). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, *1111*(1), 134-142.
- Vatakis, A., & Spence, C. (2006b). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience letters*, *393*(1), 40-44.
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & psychophysics*, *69*(5), 744-756.
- Vatakis, A., & Spence, C. (2008). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta psychologica*, *127*(1), 12-23.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871-884.
- Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive brain research*, *22*(1), 32-35.
- Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition*, *118*(1), 75-83.
- Vroomen, J., van Linden, S., De Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572-577.
- Wallace, M. T. (2004). The development of multisensory processes. *Cognitive Processing*, *5*(2), 69-83.
- Wallace, M. T., Carriere, B. N., Perrault, T. J., Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *The Journal of Neuroscience*, *26*(46), 11844-11849.

- Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, 91(3), 484-488.
- Wallace, M. T., Perrault, T. J., Hairston, W. D., & Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *The Journal of Neuroscience*, 24(43), 9580-9584.
- Wallace, M. T., Roberson, G., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), 252-258.
- Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *The Journal of Neuroscience*, 17(7), 2429-2444.
- Wallace, M. T., & Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *The Journal of Neuroscience*, 21(22), 8886-8894.
- Wallace, M. T., & Stein, B. E. (2007). Early experience determines how the senses will interact. *Journal of neurophysiology*, 97(1), 921-926.
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64, 105-123.
- Watanabe, T., Náñez, J. E., 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. *Nature neuroscience*, 5(10), 1003-1009.
- Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413(6858), 844-848.
- Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: toward a comprehensive theory. *Annual review of psychology*, 66, 197.

- Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology*, *18*(24), 1922-1926.
- Xu, J., Yu, L., Stanford, T. R., Rowland, B. A., & Stein, B. E. (2015). What does a neuron learn from multisensory experience? *Journal of neurophysiology*, *113*(3), 883-889.  
doi:10.1152/jn.00284.2014
- Yarrow, K., Minaei, S., & Arnold, D. H. (2015). A model-based comparison of three theories of audiovisual temporal recalibration. *Cognitive psychology*, *83*, 54-76.
- Yotsumoto, Y., Watanabe, T., & Sasaki, Y. (2008). Different dynamics of performance and brain activation in the time course of perceptual learning. *Neuron*, *57*(6), 827-833.
- Yu, L., Rowland, B. A., & Stein, B. E. (2010). Initiating the development of multisensory integration by manipulating sensory experience. *The Journal of Neuroscience*, *30*(14), 4904-4913.
- Yu, L., Rowland, B. A., Xu, J., & Stein, B. E. (2013). Multisensory plasticity in adulthood: cross-modal experience enhances neuronal excitability and exposes silent inputs. *Journal of neurophysiology*, *109*(2), 464-474.
- Yu, L., Stein, B. E., & Rowland, B. A. (2009). Adult plasticity in multisensory neurons: short-term experience-dependent changes in the superior colliculus. *The Journal of Neuroscience*, *29*(50), 15910-15922.
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & psychophysics*, *67*(3), 531-544.
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*(2), 198-210.
- Zilber, N., Ciuciu, P., Gramfort, A., Azizi, L., & Van Wassenhove, V. (2014). Supramodal processing optimizes visual perceptual learning and plasticity. *NeuroImage*, *93*, 32-46.

Zweig, L. J., Suzuki, S., & Grabowecky, M. (2015). Learned face-voice pairings facilitate visual search. *Psychonomic Bulletin & Review*, 22(2), 429-436. doi:10.3758/s13423-014-0685-

3

## Chapter II

### The Effects of Feedback Type and Reliability on Multisensory Perceptual Learning

#### Abstract

Recently, several studies have reported that the temporal interval over which multisensory stimuli are integrated is highly plastic in adults. However, numerous questions remain about the nature of multisensory temporal plasticity. For example, it remains unclear if changes in multisensory temporal acuity following perceptual training are dependent upon feedback. It is also not known whether changes in temporal acuity can be further enhanced by more rewarding forms of feedback. As perceptual training has been suggested as therapeutic tool to enhance temporal acuity for clinical populations, it is necessary to fully understand how feedback might elicit these changes. Here, we first sought to determine if decreasing the response-feedback contingency by presenting less reliable or unreliable feedback would impair perceptual learning for low-level audiovisual stimuli (flashes and beeps). Second, we sought to determine if alternative forms of feedback would differentially enhance multisensory temporal acuity. We observe that decreasing feedback reliability impairs perceptual learning, suggesting that multisensory perceptual learning is dependent upon the presence of a feedback signal. We also report that several different forms of feedback were able to elicit multisensory perceptual learning. Collectively, this work sheds important light on the role of feedback in driving plastic changes in multisensory temporal representations.

#### Introduction

The synthesis of information across the senses, termed multisensory integration, is essential for the development of a unified perceptual experience (Stein et al., 2010; Stein, Stanford, & Rowland, 2014). Through multisensory integration, cues from differing sensory

modalities can dramatically influence behavior, resulting in behavioral gains (e.g. improvements in sensory localization(Wallace et al., 2004) or speeded reaction times(Diederich & Colonius, 2004)) and under certain circumstances resulting in novel percepts or illusion (e.g. McGurk illusion(McGurk & MacDonald, 1976) or sound-induced flash illusion(Diederich & Colonius, 2004)). Such benefits depend critically on the capacity to properly integrate coincident sensory cues originating from the same event. The spatial and temporal relationships between sensory signals are several of the more important physical properties of a sensory event that determine whether the signals will be integrated (Stein & Meredith, 1993). Thus, the closer two sensory signals are in space or time, the more likely that these signals will be perceptually bound and the more likely multisensory enhancement will occur.

In the temporal domain, although stimulus energies from the different senses may originate simultaneously, the difference in environmental and neural propagation of these signals must be resolved and thus a flexible temporal interval during which perceptual binding may occur is essential (Vroomen & Keetels, 2010). This epoch, during which stimuli are likely to be perceived as perceptually bound, has been termed the temporal binding window (TBW) (Diederich & Colonius, 2004; Wallace & Stevenson, 2014). The TBW, typically on the order of several hundred milliseconds, provides a useful index of temporal acuity and range over which stimulus integration is likely to occur (R. A. Stevenson & Wallace, 2013; van Wassenhove, Grant, & Poeppel, 2007). The probabilistic nature of the TBW allows for a capacity to resolve perceptual meaning from events that happen at different distances with different corresponding visual and auditory arrival times. Similarly, the TBW may adaptively allow for the ability to interpret and construe meaning from the most complex multisensory stimuli, such as speech, that exhibit a high degree of temporal (and other forms of) variability. Despite the advantages of tolerating stimulus asynchrony and variability within a temporal range, if the TBW is too large, stimuli may be integrated that should be segregated (as has been seen in several clinical populations, see below). The strength of perceptual binding has been observed to relate to the

TBW as less temporal acuity (i.e., wider windows) for multisensory stimuli correlates with decreased perceptual binding (Ryan A Stevenson et al., 2014; Ryan A. Stevenson, Zemtsov, & Wallace, 2012).

The size of the TBW is not uniform but rather has been found to be heavily task- and stimulus-dependent, highly individualistic, and extremely dynamic (Fujisaki, Shimojo, Kashino, & Nishida, 2004; R. A. Stevenson & Wallace, 2013). These findings suggest that nature of the TBW is such that it develops to allow statistical inferences to be made about the relatedness of sensory stimuli, which allows for the construction of adaptive perceptual representations of the world. Knowledge of these statistical sensory relationships appears to be constructed from experience as during the course of typical development the TBW appears to narrow from infancy (Lewkowicz, 1996) to childhood and through adolescence (Hillock, Powers, & Wallace, 2011; Lewkowicz, 2014). The TBW also remains malleable in adulthood as it adjusts to represent changing stimulus relationships as repeated exposure to asynchronous audiovisual stimuli results in a shift in the window in the direction of repeated exposure (Fujisaki et al., 2004; Navarra et al., 2005; Vatakis, Navarra, Soto-Faraco, & Spence, 2007).

Recently, a number of reports have observed that the temporal acuity for multisensory stimuli remains malleable in adulthood and can be altered by perceptual learning in which explicit feedback was given during training sessions (Powers, Hevey, & Wallace, 2012; Powers, Hillock, & Wallace, 2009; Setti et al., 2014; Ryan A. Stevenson, Wilson, Powers, & Wallace, 2013). Most importantly, these studies demonstrated that changes in temporal acuity resulting from perceptual training did not occur as a result of a change in criterion, practice effects, or passive recalibration but were directly tied to the feedback individuals received regarding temporal judgments. As altered temporal acuity for multisensory stimuli has been associated with conditions such as autism spectrum disorder (ASD) and dyslexia, the prospect of improving multisensory temporal acuity through perceptual learning is of great potential clinical relevance (Ryan A Stevenson et al., 2014).

The current study sought to better elucidate the relationship between the nature of the feedback given during multisensory perceptual training and the resulting changes in temporal acuity. Unlike visual perceptual learning, the contingencies for which multisensory perceptual learning may occur have yet to be explored despite the hypothesized benefits of multisensory learning (Shams & Seitz, 2008). In the first experiment, we sought to determine if decreasing the reliability of perceptual feedback (in which the veridical relationship of the feedback to response was reduced) prevented multisensory perceptual learning. Prior investigations of visual perceptual learning have observed that uncorrelated feedback presented during perceptual training prevents multisensory learning (Herzog & Fahle, 1997) although no study has investigated if multisensory learning is dependent upon correlated feedback. If multisensory perceptual learning is dependent upon reliable feedback, we hypothesized that we would observe weaker or no changes in temporal acuity following training with less reliable feedback signals.

In the second experiment, we investigated if multisensory perceptual learning would occur when different forms of feedback likely to have varying degrees of reward were employed. We compared changes in temporal acuity following perceptual training with forms of feedback that included social (e.g., icons of happy and sad faces), symbolic (e.g., check and an X), and monetary (participants received \$ symbols following correct responses and that was linked to additional compensation) cues. Prior investigations of multisensory perceptual learning have been limited in the form of feedback available to participants during perceptual training. Furthermore, others have also observed that some forms of multisensory temporal perception are differentially modulated by the type perceptual feedback provided to study participants (Rosenthal, Shimojo, & Shams, 2009). Thus, we hypothesized that additional forms of feedback, such as those with greater reward value, might differentially enhance multisensory perceptual learning.

## **Material and methods**

### *Subjects*

Written informed consent was obtained from all study participants prior to participating in the study. A total of N = 93 individuals participated in the study with n = 45 (Age, M = 20.0 years, Gender, 23 female) individuals participating in Experiment 1 while n = 48 (Age, M = 20.1 years, Gender, 26 female) individuals participated in Experiment 2. All participants had self-reported normal hearing and normal or corrected to normal vision. All recruitment and experimental procedures were approved by the Vanderbilt University Institutional Review Board and in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### *Simultaneity Judgment (SJ) Assessment*

Participants were seated in a light and sound attenuating WhisperRoom™ (SE 2000 Series, Whisper Room Inc.) room for all tasks. Visual stimuli were presented at approximately 60 cm from the participants, who utilized a chin-rest to ensure they maintained the same distance from the computer screen. A fixation marker (1cm x 1cm) on a black background was present on the screen both between trials and throughout the duration of a trial including presentation of the visual stimulus. Participants were asked to maintain fixation on the fixation marker throughout the experiment. For the SJ task, participants were instructed to judge whether the visual stimulus and auditory stimulus “were synchronous, at the same time” or “were asynchronous, at different times” by pressing either 1 or 2 respectively using a keyboard (see Figure 2-1). The visual stimulus consisted of a white ring on a black background that subtended 7.2° of visual space with an outer diameter of 6.0 cm and an inner diameter of 3.0 cm. Visual stimuli were presented for 8.3 ms (the duration of a single screen refresh cycle) on a monitor (Samsung syncmaster 22 inch 2233 RZ LCD) with a refresh-rate of 120 Hz. The auditory stimulus consisted of a 1800 Hz tone that was presented binaurally via headphones

(Sennehiser HD 558) with no interaural time or level differences. Auditory stimuli were at 83 dB and were calibrated using a sound level meter (Larson Davis SoundTrack® LxT2). For each trial, visual and auditory stimuli were presented at true synchrony (0 ms of asynchrony) and at stimulus onset asynchronies (SOAs) of -400 and -300 to -50 ms by steps of 50 ms, for which the visual stimulus was leading, and SOAs of 50 to 300 ms by steps of 50 ms and 400 ms, for which the auditory stimulus was leading. SOAs were verified externally using an oscilloscope. Following presentation of visual and auditory stimuli, a response screen was presented at which time subjects could make a response. The inter trial interval (ITI) between trials was randomly jittered from 500 to 1500 ms. MATLAB (The MathWorks, Inc.) with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) was used to present the SJ task and all computer tasks in Experiments 1 and 2.

The SJ assessment without feedback was utilized to assess temporal acuity prior to and following training or exposure sessions for Experiments 1 and 2. For pre-training and post-training assessments, each SOA was randomly presented 20 times (total of 300 trials) for each assessment.

## **Experiment 1**

### *Protocol*

Individuals participating in Experiment 1 ( $n = 45$ ) were randomly assigned to one of three groups for which the reliability of the feedback signal was varied during the SJ training sessions. Thus, some participants completing the tasks in Experiment 1 received erroneous or false feedback during the training sessions. Training sessions were divided into 6 blocks. Prior to and following the training blocks, participants completed a SJ assessment to measure temporal acuity for the auditory and visual stimuli. All tasks for Experiment 1 were completed on the same day.

### *Simultaneity Judgment (SJ) Training Session*

The SJ training tasks asked participants to make the same judgments as the SJ assessment task, however, training tasks differed from the SJ assessment in that participants received trial-by-trial visual feedback following responses and the SOAs utilized during training sessions differed from those SOAs presented for SJ assessment. The visual and auditory stimuli were the same as used for the SJ assessment. The visual and the auditory stimulus were presented at SOAs between -150 and 150 ms by 50 ms increments such that there were 6 asynchronous SOA and true synchrony presented. Additionally, the number of trials per SOAs were not equally distributed during the training sessions. The veridical simultaneous condition had a 6:1 ratio to the asynchronous conditions such that the total number of simultaneous trials presented was equal to the total number of asynchronous trials presented. The number of synchronous and asynchronous trials were equally presented to mitigate concerns about response bias.

Visual feedback was presented for 500 ms immediately following the participant's response for each trial. Only one form of feedback (symbolic feedback, a blue-green check mark and red X) was presented for Experiment 1 and the veridical relationship of the feedback to response was altered for two groups of participants such that reliability of the feedback signal was decreased. Reliable feedback was defined as presentation of the correct feedback for each SOA-response pair such that a response of synchronous for a trial in which the SOA presented was 0 ms yielded the presentation of a blue-green indicative of the correctness of the participant's response. Thus, false or erroneous feedback was defined the presentation of the incorrect feedback for each SOA-response pair (i.e. a response of synchronous for a trial in which the SOA presented was 0 ms yielded the presentation of a red X, the exact opposite of the reliable feedback). Participants were assigned to one of three groups that received either 100% reliable feedback (n = 15), 80% reliable feedback (n = 15), or 50% reliable feedback (n = 15) during the perceptual training blocks. The veridical nature of the feedback (reliable and false

feedback) was distributed at a ratio equal to that of the SOA occurrence during the training blocks. For example, for the group receiving 50% reliable feedback, a single training block (total of 120 trials) would have false feedback on 30 asynchronous trials (5 trials with false feedback for each asynchronous SOA) and false feedback on 30 synchronous trials (60 total synchronous trials). A total of 6 training blocks (120 trials pre block, 720 total trials) comprised the training session completed by each participant.

## **Experiment 2**

### *Protocol*

A separate set of individuals participating in Experiment 2 ( $n = 48$ ) were randomly assigned to complete either SJ training with one of three different forms of feedback (training session) or complete a detection task while being exposed to auditory and visual stimuli at the SOAs as those participating in training sessions (exposure session). Training and exposure sessions were divided into 6 blocks. Prior to and following either training or exposure sessions, participants completed a SJ assessment to measure temporal acuity for the auditory and visual stimuli. Participants completed the same SJ assessment prior to and following the training session to assess temporal acuity as in Experiment 1. All tasks in Experiment 2 were completed in the same day.

### *Simultaneity Judgment (SJ) Training Session*

The SJ training task employed in Experiment 2 was identical to the SJ training paradigm employed by Experiment 1 except that the type of feedback was varied for the three different experimental groups and no participants received false feedback. Participants received one of three forms of visual feedback: symbolic feedback ( $n = 12$ ) in the form of a blue-green check mark or red X following correct and incorrect responses respectively, social feedback ( $n = 12$ ) in the form of a yellow, happy face cartoon or blue, sad face cartoon following correct and

incorrect responses respectively, or monetary feedback (n = 12) in the form of \$ signs that participants were instructed that represented additional compensation they would earn that only followed correct responses (Figure 2-1A). Each SJ training block consisted of 120 trials with a total of 720 trials presented during the training session.

### *Exposure Session*

To control for stimulus exposure, a separate group of participants (n = 12) completed an exposure session between pre-training and post-training SJ assessments. The exposure session was the same as the training session completed by other participants except that subjects completed a visual detection task rather than an SJ task. The visual and auditory stimuli presented during the exposure session were identical to the training session except that 10% of the visual stimuli were red rings. Participants in the exposure session were instructed to press 2 if they saw a white colored ring and press 1 on keyboard if they saw a red colored ring in order to maintain attention during the exposure session. The same SOAs between the auditory and visual stimuli for the training session were used during the exposure session.

## **Analysis**

### *Grand mean SOA analysis*

To determine how the overall group probability of report of synchrony changed after either training sessions or exposure sessions, individual means at each SOA were averaged to produce grand averages. A repeated-measures ANOVA was conducted for each group for which analysis of simple effects was conducted if a significant training status x SOA interaction existed. *Post hoc* pairwise comparisons of simple effects were conducted with Bonferroni correction to determine if training or exposure sessions significantly altered the report of synchrony at each SOA. Degrees of freedom were adjusted for all interactions where sphericity was violated using a Greenhouse-Geisser correction.

### *Estimation of the TBW*

Individual subjects' data on pre-training and post-training SJ assessment was fit with two sigmoid curves using the *glmfit* function in MATLAB to separately fit the data for the visual leading auditory (visual-leading TBW; also referred to as the left TBW) and auditory leading visual (auditory-leading TBW; also referred to as the right TBW) conditions (2-1B). Each participant's visual-leading, auditory-leading, and total TBW were then estimated, relative to the participant's PSS, at the SOA at which the best-fit sigmoids *y* value equaled a criterion of 70 % report of synchrony (see Stevenson & Wallace, 2013 for a detailed description of TBW estimation). The whole TBW was calculated as the sum of the auditory-leading and visual-leading TBW. Group TBWs were calculated by taking the mean of the auditory-leading, visual-leading, and total TBW for each participant. Within each group, an analysis of the effect of training was determined by paired samples t-test comparing the pre- and post-training TBW consistent with previous analyses of perceptual training related effects (Powers et al., 2009). To determine if the magnitude of the change in the TBW differed between groups, we performed group-level analyses. Group-level analysis of differences in the TBW following training or exposure sessions was conducted by performing an ANCOVA with the TBW prior to training or exposures sessions as the covariate to correct for differences in the TBW between groups following the initial SJ assessment. *Post hoc* pairwise comparisons were conducted using a Bonferroni correction to determine if the TBW was significantly altered following training or exposure sessions.

### **Results: Experiment 1**

#### *Decreasing the reliability of feedback impairs multisensory perceptual learning*

For the group receiving feedback that was 100% reliable on a trial-by-trial basis, a decrease in the probability of making simultaneity judgment at objectively asynchronous SOAs

was observed (Figure 2-2A). The introduction of less reliable feedback (i.e., 80% reliable or 50% reliable) during perceptual training sessions did not significantly alter the probability of simultaneity judgments for either group (Figure 2-2, B and C). A SOA x pre-/post-training status within-subjects repeated measures ANOVA was conducted for each group. A significant interaction was observed for only the group receiving 100% veridical feedback ( $F_{6,136,85.901} = 4.487, p < .001$ ), but not for the groups receiving 80% veridical feedback ( $F_{4,652,65.132} = 2.238, p > .05$ ) and 50% veridical feedback ( $F_{5,416,75.827} = 0.800, p > .05$ ). *Post hoc* comparisons were conducted for the simple effect of pre-/post-training status for the 100% veridical feedback group which determined that the probability of simultaneity judgment decreased at SOAs of -150 ms ( $p = .036$ ), -100 ms ( $p = .003$ ), -50 ms ( $p = .005$ ), 0 ms ( $p = .032$ ), 100 ms ( $p = .017$ ), 150 ms ( $p = .002$ ), 200 ms ( $p = .001$ ), 250 ms ( $p = .005$ ), and 300 ms ( $p = .001$ ).

Next, we examined how feedback reliability altered the TBW following perceptual training. We first determined if perceptual training altered the TBW within each group by conducting paired samples t-tests for the whole, auditory-leading leading, and visual-leading TBWs pre- and post-training. We observe that the whole TBW is significantly smaller following perceptual training for the 100% veridical feedback group ( $p = .001$ ), but not for the 80% ( $p > .05$ ) or 50% ( $p > .05$ ) reliable feedback groups (2-3). The auditory-leading ( $p = .022$ ) and visual-leading ( $p < .001$ ) TBWs for the 100% veridical feedback groups were also significantly narrower following training, whereas no change in the auditory-leading or visual-leading TBWs were observed for the groups receiving 80% or 50% reliable feedback.

An ANCOVA using the pre-training TBW as the covariate was utilized to determine if altering the reliability of feedback during perceptual training sessions differentially altered the TBW post-training across groups. By considering the pre-training TBW as a covariate, we were better able to adjust for any effects that might arise due to pre-training differences in temporal acuity between the groups and better assess the effects of training. The post-training TBW was determined to be differentially altered by training group ( $F_{2,41} = 3.560, p = .038$ ). *Post hoc*

comparisons determined that only the post-training whole TBW for the group receiving 100% reliable feedback was smaller than the TBW for the group receiving 50% reliable feedback, ( $p = .040$ ). The visual-leading ( $F_{2,41} = 2.699$ ,  $p > .05$ ) and auditory-leading TBWs ( $F_{2,41} = 1.028$ ,  $p > .05$ ), when analyzed separately, were not significantly altered by feedback when analyzed at an individual level.

## **Results: Experiment 2**

### *Perceptual Training with Symbolic, Social, and Monetary Feedback on a SJ task Enhances Multisensory Temporal Acuity*

Following perceptual training sessions, an immediate decrease in the probability of simultaneity judgment at SOAs that were objectively asynchronous was observed for all subjects receiving feedback during perceptual training sessions (Figure 2-4). For each group receiving feedback, a SOA x pre-/post-training status within subjects factors repeated measures ANOVA revealed a significant interaction: symbolic feedback ( $F_{(14,154)} = 2.91$ ,  $p = .001$ ), social feedback ( $F_{(14,154)} = 4.58$ ,  $p < .001$ ), and monetary feedback ( $F_{(14,154)} = 2.20$ ,  $p = .010$ ). *Post hoc* analysis of the simple effect for pre-/post-training status was conducted with Bonferroni correction for each group receiving feedback and revealed significant decreases in the probability of reporting simultaneity at the following individual SOAs for each group: symbolic feedback at -50 ms ( $p = .043$ ), 0 ms ( $p = .013$ ), 100 ms ( $p = .025$ ), 150 ms ( $p = .005$ ), 200 ms ( $p = .031$ ), and 250 ms ( $p = .017$ ); social feedback at 50 ms ( $p = .020$ ), 150 ms ( $p = .010$ ), 200 ms ( $p = .017$ ), 250 ms ( $p = .023$ ), 300 ms ( $p = .030$ ), and 400 ms ( $p = .015$ ); and monetary feedback at -50 ms ( $p = .039$ ), 150 ms ( $p = .025$ ), 200 ms ( $p = .001$ ), and 400 ms ( $p = .013$ ). The majority of these changes in simultaneity judgment occurred for SOAs for conditions in which the visual stimulus was leading the auditory stimulus.

In contrast to individuals receiving feedback during perceptual training sessions, passive exposure to asynchronous stimuli during the exposure sessions did not affect the probability of

simultaneity judgment at any of the tested SOAs (Figure 2-2). A SOA x pre-/post-exposure status within subjects factors repeated measures ANOVA did not reveal a significant interaction ( $F_{(14,154)} = 1.648, p > .05$ ).

We next determined if perceptual training sessions altered the TBW within each group using paired-samples t-tests for the whole TBW, as well as for the auditory-leading and the visual-leading sides of the distribution (Figure 2-5). The whole TBW was significantly smaller following perceptual training for groups receiving symbolic ( $p = .006$ ), social ( $p = .020$ ), and monetary ( $p = .016$ ) feedback, but not for individuals passively exposed to asynchronous stimuli during the perceptual training sessions ( $p > .05$ ). Analysis of the individual TBWs revealed that for all groups receiving feedback, the visual-leading side of the TBW distribution was significantly smaller following perceptual training (symbolic,  $p = .022$ ; social,  $p = .002$ ; monetary,  $p = .015$ ). In contrast, the auditory-leading side of the TBW distribution was only smaller for the group receiving symbolic feedback ( $p = .014$ ) compared to the groups receiving social ( $p > .05$ ) or monetary feedback ( $p > .05$ ). No changes in either the auditory-leading ( $p > .05$ ) or visual-leading ( $p > .05$ ) TBWs were observed for the group passively exposed to asynchronous stimuli.

An ANCOVA using the TBW prior to training as the covariate determined that the post-training TBW was differentially altered following training and exposure blocks ( $F_{3,43} = 7.920, p < .001$ ). *Post hoc* comparisons determined that a significant decrease in the total TBW was observed for all groups receiving feedback during training sessions relative to those participants who were passively exposed to asynchronous stimuli. The TBW was not significantly affected by the type of feedback utilized during perceptual training as the post-training TBWs for the groups receiving feedback (symbolic, monetary, or social) were not significantly different. Thus, while temporal acuity was enhanced for all groups receiving feedback relative to the exposure group, there was no difference in the capacity for any single form to enhance temporal acuity. We subsequently sought to determine if the auditory- and visual-leading TBWs were also differentially post training and exposure blocks. Separate ANCOVAs determined that the post-

training TBW was significantly affected by training or exposure blocks for both the visual-leading TBW ( $F_{3,43} = 6.078$ ,  $p = .002$ ) and auditory-leading TBW ( $F_{3,43} = 4.756$ ,  $p = .006$ ). Only the visual-leading TBW was significantly smaller for all groups receiving feedback (symbolic, social, and monetary feedback) in comparison to the exposure group. The auditory-leading TBW was only significantly smaller than the exposure group for the group receiving symbolic feedback.

### **Initial Size of TBW Predicts Training Related Changes**

Prior investigations had observed that the initial measurement of the TBW pre-training predicted the degree of change in the TBW following perceptual training; thus, we choose to examine if the initial TBW was predictive of training related changes in both Experiments 1 and 2. As the group receiving symbolic feedback in Experiment 1 and the 100% veridical feedback group in Experiment 2 completed identical tasks, these data sets were combined into a single group. The initial TBW was correlated with the change in the TBW determined by the difference of the post-training TBW – pre-training TBW.

We observe that the size of the TBW pre-training was highly predictive of the degree to which the TBW changed following perceptual training for all groups receiving 100% reliable feedback (Figure 2-6). Thus, in this group, the larger the individual TBW, the greater the degree that the TBW narrowed: symbolic feedback ( $r = -.508$ ,  $n = 27$ ,  $p = .007$ ), social feedback ( $r = -.781$ ,  $n = 12$ ,  $p = .003$ ), and monetary feedback ( $r = -.805$ ,  $n = 12$ ,  $p = .002$ ). In comparison, the size of the TBW pre-training was not predictive of changes in the TBW following passive exposure to the same stimuli ( $r = .495$ ,  $n = 12$ ,  $p = .102$ ). Interestingly, for the groups received less (i.e., 80%) or no (i.e., 50%) reliable feedback, only in the 50% reliable feedback group was the pre-training TBW predictive of training related changes in the TBW ( $r = -.654$ ,  $n = 15$ ,  $p = .008$ ). This paradoxical effect may have been driven by differences in the initial size of the TBW, in that for those individuals with the smallest TBW prior to training, the post-training effects

appeared to be widening of the TBW, whereas for those individuals with the largest TBW prior to training there was a narrowing effect of the training.

## **Discussion**

Multisensory training may have additional benefits over unisensory perceptual training as perceptual learning for multisensory stimuli has been observed to occur more rapidly (Kim, Seitz, & Shams, 2008) and to elicit superior perceptual gains (Barakat, Seitz, & Shams, 2015). A growing interest in the capacity for multisensory perceptual learning to enhance both unisensory and multisensory function warrants understanding the optimal training contingencies for which learning is elicited. Prior to this investigation, the capacity for and contingencies necessary for feedback to elicit multisensory learning had yet to be elucidated. In Experiment 1, we observe that diminishing the reliability of the feedback is sufficient to prevent perceptual learning at the group level. We were surprised to observe in Experiment 2 that multiple forms of feedback have equal capacity to elicit multisensory perceptual learning and that these changes arise rapidly following a single, hour-long training session. Our findings illustrate that multisensory perceptual learning is elicited only in the presence of feedback and under circumstances where the response-feedback relationship is evident.

Important to this study is the rapid nature of the changes in temporal acuity for audiovisual stimuli. Studies of perceptual learning for unisensory stimulus properties often report that significant perceptual learning only occurs over the course of many days of training (Kim et al., 2008; Aaron R. Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006). A hypothesized benefit of multisensory training is that perceptual learning will occur more rapidly and more efficiently (Shams & Seitz, 2008). Here we show that a single, hour-long session of perceptual training is sufficient to significantly enhance temporal acuity for a simple audiovisual stimulus. This finding is consistent with previous observations that the largest reduction in the TBW (i.e. greatest change in temporal acuity) occurred following the first day of perceptual training

(Powers et al., 2009). We also observe that following this single training interval the TBW narrows to approximately 200-250 ms. This value is very concordant with prior work that trained for longer intervals, suggesting that there may exist a limit to which the TBW may narrow. It is also possible that this apparent limit of the TBW is related to the task difficulty during training sessions and that increasing the difficulty of the perceptual training session task may result in further reduction in the TBW. Such task-difficulty mediated effects have been seen in prior studies of visual perceptual learning (Liu, Lu, & Doshier, 2012; Aaron R. Seitz et al., 2006).

In Experiment 1, we find that limiting the reliability of the feedback prevents multisensory perceptual learning, even if the amount of false feedback is 20%. This suggests that the presence of a reliable feedback signal is necessary to elicit multisensory learning. These findings support prior studies of unisensory perceptual learning that have also found that the magnitude of perceptual learning is dependent upon the amount of feedback (Liu, Doshier, & Lu, 2014) and that false or uncorrelated feedback impairs perceptual learning (Herzog & Fahle, 1997). Interestingly, it appears that when examining our findings at the individual level, plastic changes in the TBW may be prevented only for the group receiving 80% reliable feedback. Very little change in the TBW was observed for all individuals in the 80% reliable feedback group. In contrast, most individuals in the 50% reliable feedback group showed changes, although the direction of these changes (narrowing versus widening) appeared dependent upon initial (i.e., pre-training) window size (See Figure 2-6). Finally, we observe that the pre-training TBW is predictive of the change in the TBW post-training for both groups receiving 100% and 50% reliable feedback. In contrast, the pre-training TBW is not predictive of training related change in the TBW for the 80% reliable feedback group. We speculate that the initial TBW remained predictive of the training related change in the TBW for individuals in the 50% reliable feedback group because this group may have more readily perceived that false feedback was presented on some trials. Thus, for individual participants with wide TBWs in the 50% reliable feedback group, the distinction between false feedback may be apparent and thus these participants may

still attempt to utilize some aspect of the feedback signal during perceptual training. Similarly, for individual participants with narrow TBWs in the 50% reliable feedback group, if some aspect of the false feedback signal was taken into account, these individuals would be expected to widen.

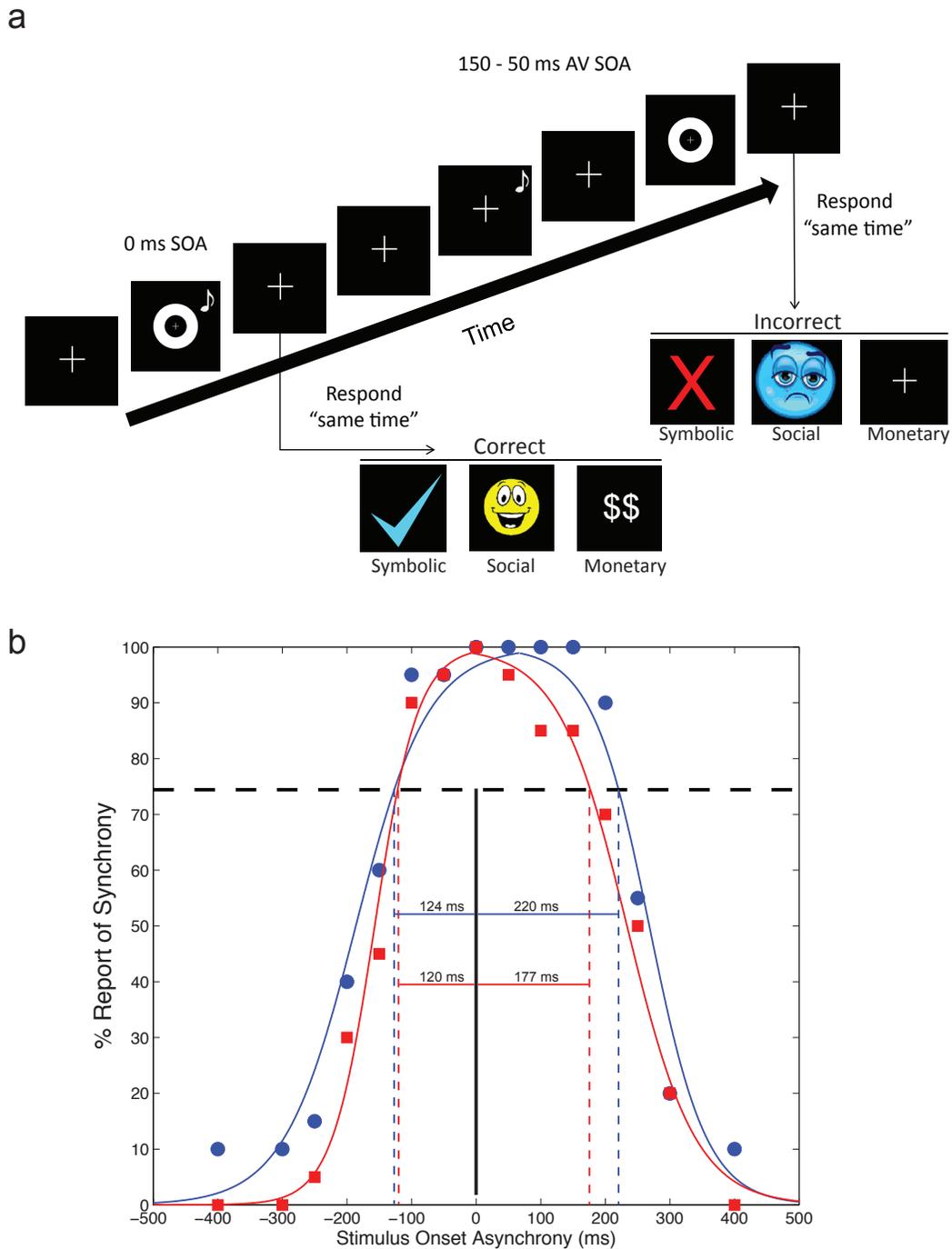
The presence of feedback appears to be sufficient to elicit multisensory perceptual learning independent of the rewarding nature for the types of feedback in this study. These findings suggest that feedback in this paradigm engages a common set of neural structures. Indeed, maintaining response-reward relationship is sufficient to elicit visual perception in the absence of stimulus awareness or explicit knowledge of the reward contingencies, suggesting that engaging reward systems is the critical gating factor for perceptual learning rather than attention or other top-down cognitive processes (Aaron R Seitz, Kim, & Watanabe, 2009). Furthermore, evidence suggests that two of the reward types utilized by this study, monetary and social rewards, engage much of the same neural circuitry, particularly the striatum (Izuma, Saito, & Sadato, 2008). While social and nonsocial rewards also likely engage similar reward-related neural structures, it is hypothesized that these reward types engage some distinct aspects of the neural circuitry for reward although further investigation is needed to elucidate these differences (Ruff & Fehr, 2014). Although the symbolic feedback (green check mark; red X) utilized in our study is inherently a social construct, we interpret this form of feedback to be the most socially neutral. Although others have observed that another measure of multisensory temporal acuity, the sound-induced flash illusion, was resistant to a more neutral form of feedback ("right" in green font, "wrong" in red font), whereas monetary based feedback decreased perception of the illusion, we do not observe a similar differentiation (Rosenthal et al., 2009). Behavioral evidence from one study supports this distinction as social rewards (happy or angry facial expressions) and nonsocial rewards (red and green lights) differentially affect learning when utilized as reinforcement signals (Hurlemann et al., 2010). As symbolic feedback in this study elicited equivalent enhancement of temporal acuity as other more social forms of

feedback, this suggests that symbolic feedback could be utilized as an equally effective feedback signal for individuals with ASD for which social stimuli are adverse (Delmonte et al., 2012; Stavropoulos & Carver, 2014). Perceptual training has been suggested to have therapeutic applications in ASD as individuals with ASD have wider TBW than typically developed peers (Ryan A Stevenson et al., 2014; Wallace & Stevenson, 2014). It is hypothesized that if perceptual training is capable of enhancing temporal acuity in ASD, the resulting enhancement in temporal acuity could lead to overall greater audiovisual integration and that may cascade into enhanced speech perception and improve social communication (given that both of these domains are dependent upon effective audiovisual integration).

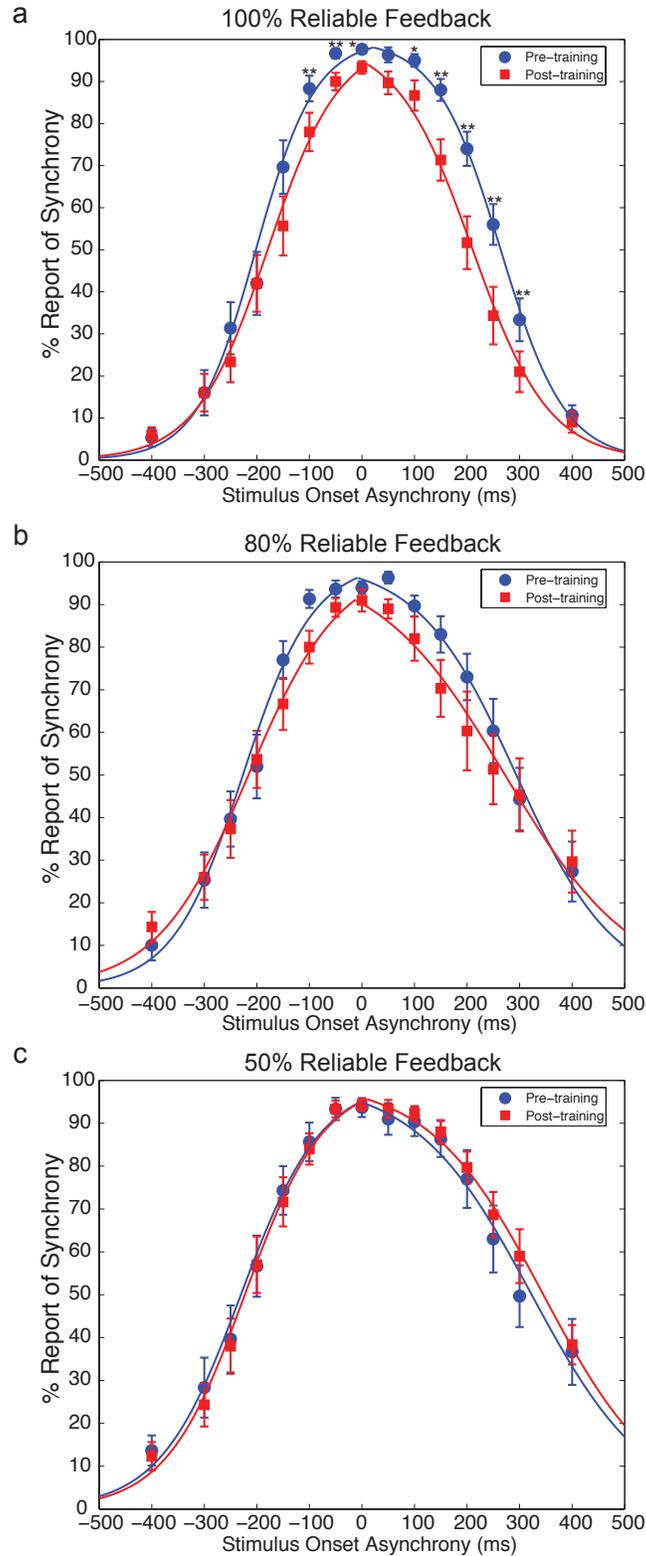
This study illustrates that multisensory perceptual learning occurs rapidly when individuals have access to feedback regarding perceptual responses. Degrading the reliability of the feedback signal, even on as few as 20% of the trials, is sufficient to prevent multisensory perceptual learning. Additionally, our findings show that the mere presence of a reinforcement signal, even a relatively neutral symbolic signal such as a check mark, is sufficient to engender multisensory temporal plasticity. Such a finding has interesting clinical and therapeutic implications in the context of autism, where deficits in multisensory temporal perception have been established, and where reward signals with social value (i.e., such as smiling and frowning faces) may be far from optimal. Hence, training approaches founded in more symbolic level rewards may be a better strategy for shaping multisensory temporal perception, and the higher order domains (i.e., social communication) that are dependent upon audiovisual temporal function.

## **Acknowledgements**

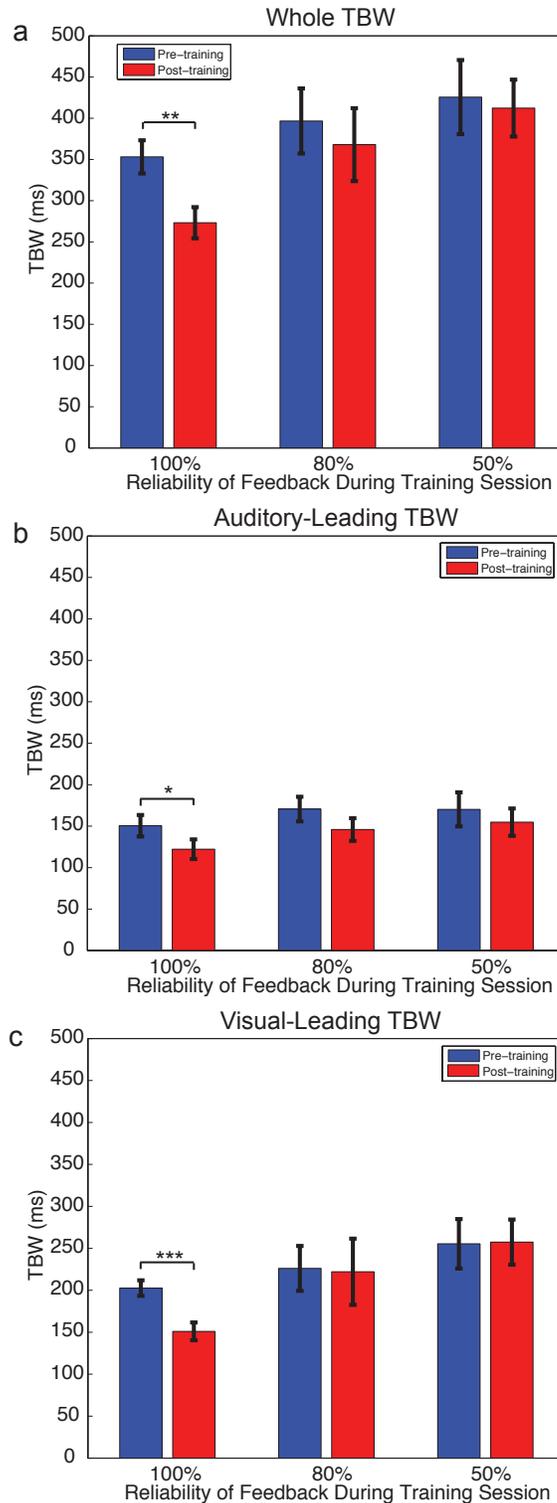
This work was supported by NIH grants CA183492 and HD083211.



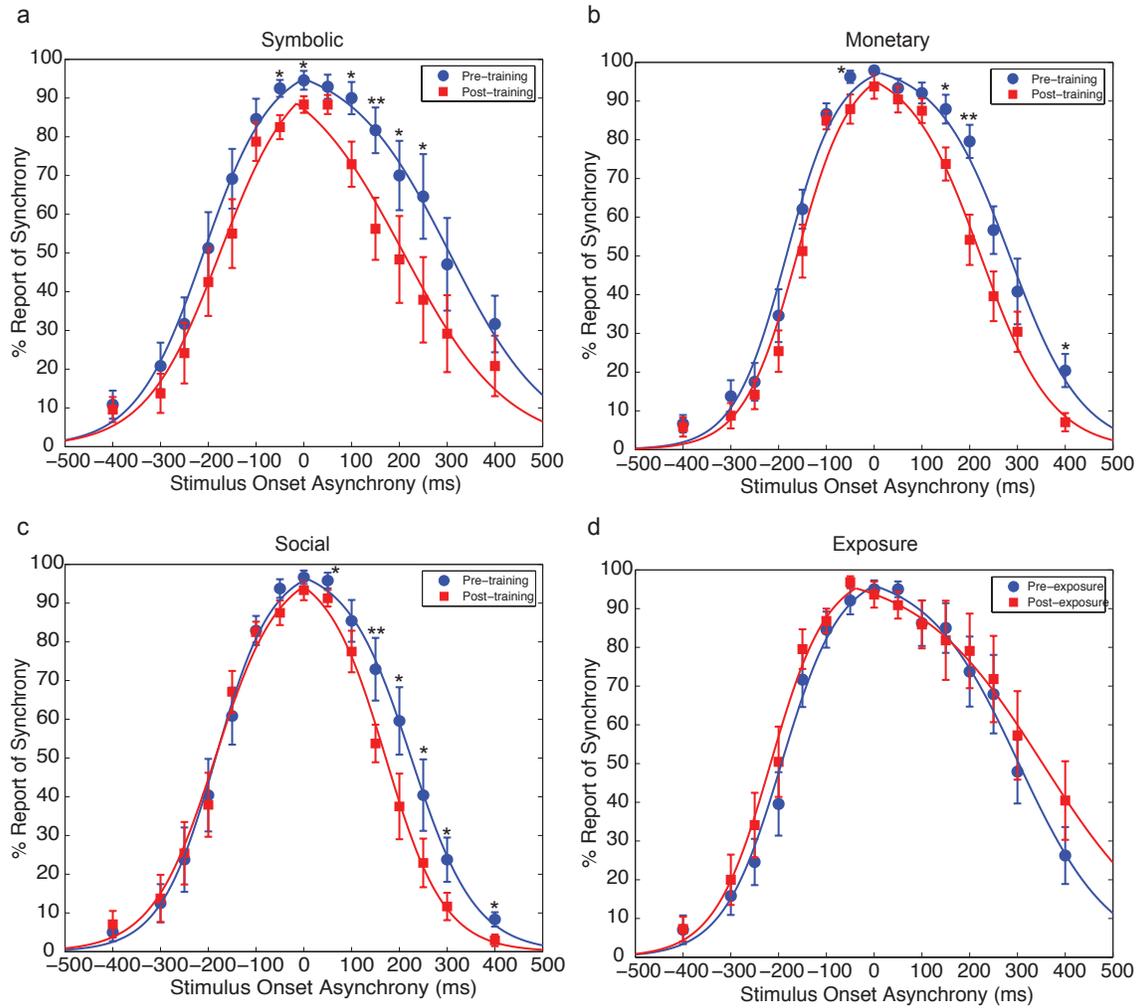
**Figure 2-1. Study Methods** a) Participants completed a simultaneity judgment (SJ) task that asked them to judge whether a “flash” (visual) and “beep” (auditory) occurred synchronously, at the same time, or asynchronously, at different times. During perceptual training sessions, feedback followed responses to indicate the correctness of the individual’s response. For the social feedback, correct responses were followed by a yellow cartoon of a happy face while incorrect responses were followed by a blue cartoon image of a sad face (actual images not shown due to copyright). b) Data from representative participant who completed the SJ task prior to perceptual training (blue circles) and following perceptual training (red squares). Sigmoid curves fit to this participant’s data are shown prior to (blue lines) and following (red lines) perceptual training. The TBW was measured by determining the SOAs at which the subject reported synchrony on greater than 75% of trials.



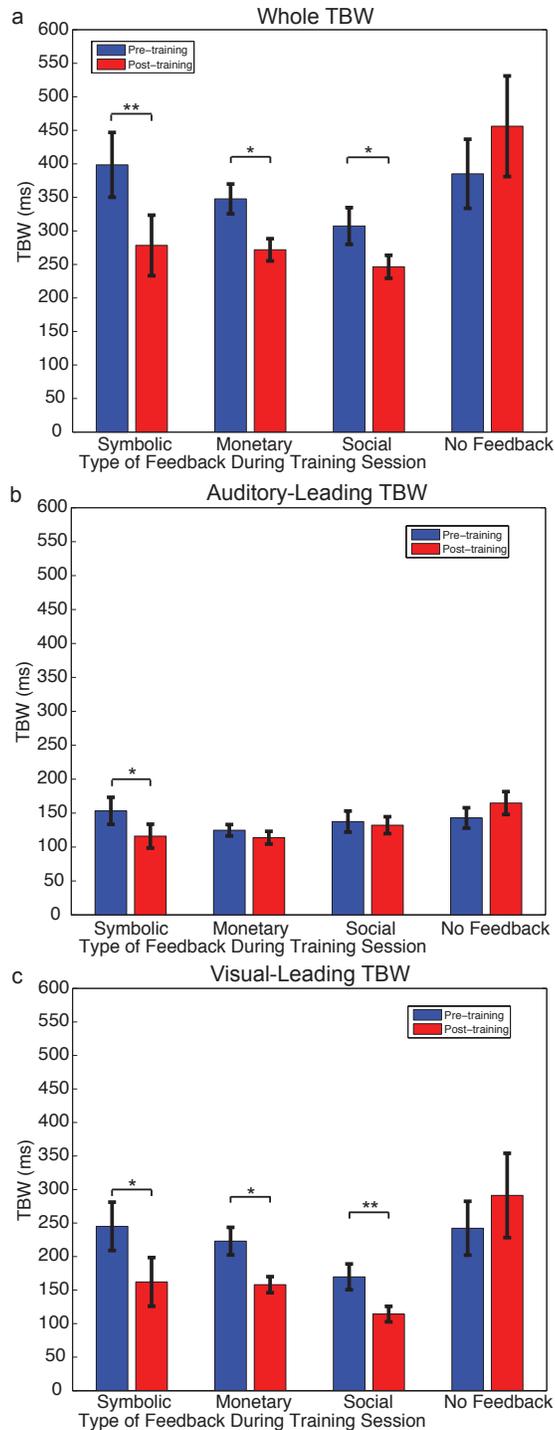
**Figure 2-2. Only Reliable Feedback Alters the Probability of Synchrony Judgment** a) Feedback decreases the probability of reporting synchrony for participants receiving 100% reliable feedback b) 80% and c) 50% reliable feedback during perceptual training does not alter the probability of reporting synchrony. Error bars,  $\pm 1$  SEM. (\*,  $p < .05$ ; \*\*,  $p < .01$ ).



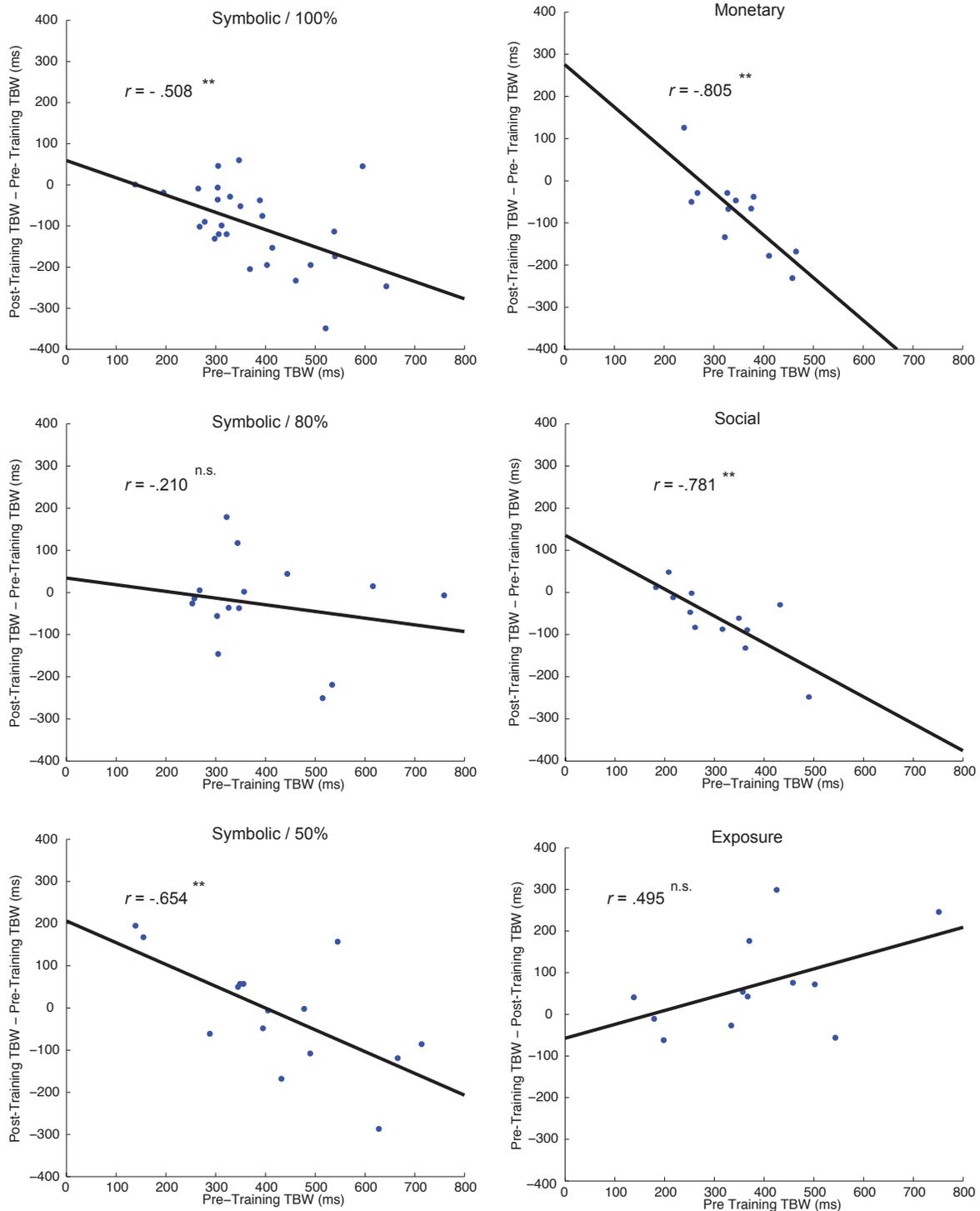
**Figure 2-3. Decreasing Feedback Reliability Impairs Multisensory Perceptual Learning** Narrowing of the whole (a), auditory-leading (b), and visual-leading (c) TBWs was observed for the group receiving 100% feedback during perceptual training. No significant change in either the whole, auditory-leading, or visual-leading TBWs was observed for the groups receiving either 80% or 50% reliable feedback. Error bars,  $\pm 1$  SEM, (\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ).



**Figure 2-4. Probability of Simultaneity Judgment Only Altered When Feedback is Present**  
 Feedback decreases the probability of reporting synchrony for participants receiving a) symbolic, b) monetary, or c) social feedback. d) Passive exposure to asynchronous stimuli do not alter the probability of reporting synchrony at any SOA. Error bars,  $\pm 1$  SEM. (\*,  $p < .05$ ; \*\*,  $p < .01$ ).



**Figure 2-5. Multiple Types of Feedback Elicit Multisensory Perceptual Learning** a) Narrowing of the TBW is observed following perceptual training for participants receiving symbolic, monetary, or social feedback. The whole TBW was observed to be narrower post-training TBW for all groups receiving feedback unlike those participants passively exposed to asynchronous stimuli whose TBW was not significantly altered following the passive exposure sessions. b) A significant decrease in the auditory-leading TBW was only observed for those individuals receiving symbolic feedback. c) Narrowing of the visual-leading TBW was observed for all groups receiving feedback during perceptual training while no change was observed in the visual-leading TBW following passive exposure. Error bars,  $\pm 1$  SEM. (\*,  $p < .05$ ; \*\*,  $p < .01$ ).



**Figure 2-6. Temporal Acuity Prior to Perceptual Training Predicts Training Related Changes** Initial size of TBW predicts change in TBW following training for groups receiving objective monetary and social feedback in Experiment 1 and symbolic feedback in Experiments 1 and 2 (both groups completed identical tasks). Interestingly, the initial size of the TBW also predicts the change in the TBW for the group receiving 50% erroneous feedback in Experiment 2. The initial TBW was not predictive of the change in the TBW following passive exposure in Experiment 1 or perceptual training with 20% erroneous feedback in Experiment 2.

## References

- Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, *25*(2), R60-R61.  
doi:10.1016/j.cub.2014.12.011
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433-436.
- Delmonte, S., Balsters, J. H., McGrath, J., Fitzgerald, J., Brennan, S., Fagan, A. J., & Gallagher, L. (2012). Social and monetary reward processing in autism spectrum disorders. *Mol. Autism*, *3*(7).
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388-1404.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature neuroscience*, *7*(7), 773-778.
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision research*, *37*(15), 2133-2141.
- Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*(3), 461-467.  
doi:10.1016/j.neuropsychologia.2010.11.041
- Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., . . . Maier, W. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *The Journal of Neuroscience*, *30*(14), 4999-5007.
- Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*(2), 284-294.
- Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *Plos One*, *3*(1), e1532.

- Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1094.
- Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental psychobiology*, 56(2), 292-315.
- Liu, J., Doshier, B., & Lu, Z.-L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision research*, 99, 46-56. doi:10.1016/j.visres.2014.01.001
- Liu, J., Lu, Z.-L., & Doshier, B. A. (2012). Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision research*, 61, 15-24. doi:10.1016/j.visres.2011.12.002
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices.
- 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. *Cognitive brain research*, 25(2), 499-507.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4), 437-442.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, 32(18), 6263-6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, 29(39), 12265-12274.
- Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain topography*, 21(3-4), 185-192.
- Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, 15(8), 549-562.
- Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron*, 61(5), 700-707.

- Seitz, A. R., Nanez, J. E., Sr., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, *6*(9), 966-973. doi:10.1167/6.9.9
- Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia*, *61*, 259-268.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, *12*(11), 411-417.
- Stavropoulos, K. K., & Carver, L. J. (2014). Reward anticipation and processing of social versus nonsocial stimuli in children with and without autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *55*(12), 1398-1408.
- Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., . . . Sathian, K. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*(10), 1713-1720.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*: The MIT Press.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, *15*(8), 520-535.
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*, *34*(3), 691-697.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249-261. doi:10.1007/s00221-013-3507-3

- Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research*, 225(4), 479-489. doi:10.1007/s00221-012-3387-y
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. *Journal of Experimental Psychology-Human Perception and Performance*, 38(6), 1517-1529. doi:10.1037/a0027339
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.
- Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, 181(1), 173-181.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, 72(4), 871-884.
- Wallace, M. T., Roberson, G., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), 252-258.
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64, 105-123.

## Chapter III

### Multisensory Perceptual Learning is Dependent Upon Task Difficulty

#### Abstract

There has been a growing interest in developing behavioral tasks to enhance temporal acuity as recent findings have demonstrated changes in temporal processing in a number of clinical conditions. Prior research has demonstrated that perceptual training can enhance temporal acuity both within and across different sensory modalities. Although certain forms of unisensory perceptual learning have been shown to be dependent upon task difficulty, this relationship has not been explored for multisensory learning. The present study sought to determine the effects of task difficulty on multisensory perceptual learning. Prior to and following a single training session, participants completed a simultaneity judgment (SJ) task, which required them to judge whether a visual stimulus (flash) and auditory stimulus (beep) presented in synchrony or at various stimulus onset asynchronies (SOAs) occurred synchronously or asynchronously. During the training session, participants completed the same SJ task but received feedback regarding the accuracy of their responses. Participants were randomly assigned to one of three levels of difficulty during training: easy, moderate and hard, which were distinguished based on the SOAs used during training. We report that only the most difficult (i.e., hard) training protocol enhanced temporal acuity. We conclude that perceptual training protocols for enhancing multisensory temporal acuity may be optimized by employing audiovisual stimuli for which it is difficult to discriminate temporal synchrony from asynchrony.

#### Introduction

Accurate perception of the world requires synthesizing information across multiple sensory modalities (Stein, Stanford, & Rowland, 2014). The temporal proximity of these sensory

signals is an important property for determining if these signals arise from a common event or distinct events (Vroomen & Keetels, 2010). Accordingly, the capacity to accurately utilize temporal cues to integrate sensory signals from different modalities is essential for developing perceptual coherence (Stein & Stanford, 2008; Wallace & Stevenson, 2014) and for driving the well-established behavioral benefits associated with multisensory integration (Diederich & Colonius, 2004; Sumbly & Pollack, 1954). Numerous studies have demonstrated that the integration of multisensory stimuli occurs over a range of time in which the separate sensory signals may be proximate, but asynchronous (Dixon & Spitz, 1980; Meredith, Nemitz, & Stein, 1987; Vroomen & Keetels, 2010). This epoch, which typically spans several hundred milliseconds, has been termed the temporal binding window (TBW) (Wallace & Stevenson, 2014).

The TBW has emerged as a useful construct for indexing temporal acuity for multisensory stimuli, particularly audiovisual stimuli (Wallace & Stevenson, 2014). As the TBW is probabilistic in nature, it flexibly resolves differences in the neural and environmental propagation times of sensory signals, allowing for appropriate sensory integration (Vroomen & Keetels, 2010). Furthermore, the TBW appears to be flexibly specified depending upon the nature of the stimuli to be integrated, such that narrower TBWs are reported for simple audiovisual stimuli (e.g. flashes and beeps) and wider TBWs are reported for more complex stimuli (e.g. audiovisual speech) (Dixon & Spitz, 1980; R. A. Stevenson & Wallace, 2013). It has been suggested that the TBW reflects experience with the statistical relationships of sensory stimuli as over the course of development, the TBW appears to narrow to more accurately reflect the statistics of the environment (Chen, Shore, Lewis, & Maurer, 2016; Hillock, Powers, & Wallace, 2011; Hillock - Dunn & Wallace, 2012).

Although the TBW makes a great deal of ecological sense in regards to tolerating the natural asynchronies for stimuli in our world, if the TBW is too large, as is observed for those

with autism spectrum disorder (ASD), stimuli that should be segregated will instead be integrated . A large TBW (i.e. decreased temporal acuity) has also been associated with decreased strength of overall multisensory integration (Ryan A Stevenson et al., 2014; Ryan A. Stevenson, Zemtsov, & Wallace, 2012). For example, in individuals with ASD, the size of the TBW has been reported to be negatively correlated with perceptual fusion of audiovisual speech as measured by the McGurk illusion (i.e. larger TBWs are associated with lower rates of perceptual fusion) (Ryan A Stevenson et al., 2014). As multisensory temporal acuity appears to be impaired for those with ASD, there has been a growing interest in identifying ways to improve temporal performance, with the goal that these improvements in (multi)sensory function will generalize to improvements in domains of weakness such as social communication (Wallace & Stevenson, 2014).

Recent work has demonstrated that perceptual training with feedback can produce significant and lasting changes in temporal acuity (i.e. narrowing of the TBW) for audiovisual stimuli (Powers, Hillock, & Wallace, 2009; Setti et al., 2014; Ryan A. Stevenson, Wilson, Powers, & Wallace, 2013). In addition, subsequent work has found that this improvement in audiovisual temporal acuity generalizes to other measures that are dependent upon temporal processing (Powers III, Hillock-Dunn, & Wallace, 2016). Although multisensory perceptual learning appears to exhibit similarities to modality-specific (i.e., unisensory) perceptual learning, much remains unknown as to how perceptual training enhances multisensory acuity. One aspect of perceptual training that might influence the extent of multisensory perceptual learning is task difficulty. For example, in the visual domain, the difficulty of the task appears to be a strong determinant of whether learning will or will not occur (Ahissar & Hochstein, 1997; DeLoss, Watanabe, & Andersen, 2014; Aaron R. Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006; Wang, Song, Qu, & Ding, 2010). Furthermore, based on evidence from other studies of perceptual training, it has been suggested that the greatest changes in behavioral performance are observed when the stimuli presented are sufficiently challenging while also

eliciting sufficient rewarding feedback (Mishra & Gazzaley, 2014). Often, adaptive procedures (Deveau, Ozer, & Seitz, 2014; Mishra, de Villers-Sidani, Merzenich, & Gazzaley, 2014) or baseline performance measurements (Anguera et al., 2013) are employed to control for task difficulty prior to or during the course of a training session as individual differences in perceptual abilities are likely to be present prior to perceptual training. Accordingly, such individualized perceptual training protocols that equate differences in perceptual ability allow for a clearer interpretation of the actual effects of perceptual training (Anguera et al., 2013). Furthermore, perceptual training protocols that account for an individual's baseline performance (in this case temporal acuity) and subsequently individualize a training regimen based on this baseline performance data are most likely to elicit the greatest changes in perception following training (Mishra, Anguera, & Gazzaley, 2016). As perceptual training has been suggested as a potential therapeutic approach to enhance multisensory temporal acuity, it is crucial to understand how training protocols may be optimized to enhance perceptual learning.

The present study sought to address if task difficulty affects the capacity for multisensory perceptual training to enhance temporal acuity. We employed a perceptual training paradigm similar to prior studies in which a simultaneity judgment (SJ) task was employed to measure the TBW prior to and following training. Prior studies of multisensory perceptual learning have, instead, employed a single, fixed set of stimuli that remained constant during training sessions and was independent of an individual participant's temporal acuity (Powers, Hevey, & Wallace, 2012; Powers et al., 2009; Powers III et al., 2016). Unlike previous studies, we individually tailored the stimuli presented during the training sessions based on a participant's pre-training temporal acuity. Individually tailoring the training stimuli allowed us to develop three training protocols of varying difficulty based on an individual participant's ability to detect asynchrony. Importantly, by presenting SOAs that were equated based on each individual's likeliness to perceive audiovisual asynchrony, we were able to control for individual differences in temporal acuity prior to training (in contrast, changes in temporal acuity following perceptual training

employing a single set fixed SOAs may be influenced by pre-training temporal acuity). We hypothesized that increasing task difficulty during training sessions would potentiate the effects of perceptual training.

## **Methods**

### *Subjects*

All participants received informed consent prior to participating in this study. A total of  $N = 51$  typically-developing adults were included in the analyses for this study (Age,  $M = 20.21$  years; Gender, 28 female). An additional 6 individuals were enrolled in the study but were excluded from final analyses as they did not meet the minimal performance criteria during the SJ task. Participants were excluded from final analyses if they did not perceive at least  $\geq 80\%$  of trials to be synchronous for at least one SOA across all SOAs measured. All participants had self-reported normal hearing and normal or corrected to normal vision. All recruitment and experimental procedures for this study involving human participants were approved by the Vanderbilt University Institutional Review Board and in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### *Stimuli*

All stimuli were presented in a light and sound attenuating WhisperRoom™ (SE 2000 Series, Whisper Room Inc.) room in which participants were seated approximately 60 cm from a computer monitor. The visual stimulus consisted of a white ring on a black background that subtended  $7.2^\circ$  of visual space with an outer diameter of 6.0 cm and an inner diameter of 3.0 cm. Visual stimuli were presented for 8.3 ms (the duration of a single screen refresh cycle) on a monitor (Samsung syncmaster 22 inch 2233 RZ LCD) with a refresh-rate of 120 Hz. A visual fixation marker (1cm x 1cm) on a black background was present on the screen both during the intertrial interval (ITI) and throughout the duration of a trial, which included presentation of the

visual stimulus (See Figure 3-1A). Participants were asked to maintain fixation on the fixation marker throughout the experiment. The auditory stimulus consisted of a 10 ms, 1800 Hz tone that was presented binaurally via headphones (Sennheiser HD 558) with no interaural time or level differences. Auditory stimuli were presented at 83 dB and were calibrated using a sound level meter (Larson Davis SoundTrack® LxT2).

### *Experimental Procedure*

Participants completed a simultaneity (SJ) judgment task (See Figure 3-1A) during pre-training, training, and post-training trial blocks similar to SJ tasks utilized by previous studies of multisensory perceptual learning (Powers et al., 2012; Powers et al., 2009). For each trial of the SJ task, the visual and auditory stimuli were presented either synchronously or asynchronously at various stimulus onset asynchronies (SOAs). Following presentation of visual and auditory stimuli, a response screen was presented at which time participants could make a response. Participants were instructed to judge whether the visual stimulus and auditory stimulus “were synchronous, at the same time” or “were asynchronous, at different times” by pressing either 1 or 2 respectively using a keyboard. The ITI between trials was randomly jittered from 500 to 1500 ms. MATLAB (The MathWorks, Inc.) with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) was used to present stimuli and record participant responses during the SJ task. To ensure accurate timing of the presentation of auditory and visual stimuli, SOAs were verified externally using an oscilloscope.

### *Experimental Protocol*

Participants completed all tasks over the course of a single 1.5-2.0 hour session. The SJ task without feedback was utilized to assess temporal acuity (i.e. derive the TBW) prior to and following perceptual training. For pre-training and post-training SJ assessments, each SOA ( $\pm 400, 300, 250, 200, 150, 100, 50, \text{ and } 0$  ms) was randomly presented 20 times (total of 300

trials) for each assessment. Following the completion of the pre-training SJ assessment, participants completed six perceptual training blocks consisting of 120 trials per block (720 total trials across all training blocks). SOAs presented during training blocks were determined on a subject-by-subject basis dependent upon the initial estimate of the TBW derived from the pre-training SJ assessment and the training group to which the participant was randomly assigned.

### *Perceptual Training*

The structure and stimuli of the SJ task during training blocks was identical to the pre-training and post-training tasks, however, training tasks differed from the SJ assessment in that participants received trial-by-trial visual feedback following responses. Visual feedback was presented for 500 ms immediately following the participant's response. Visual feedback consisted of a blue-green check mark or red X following correct and incorrect responses respectively (Figure 3-1B).

Participants were randomly assigned to one of three training groups for which the difficulty of the task varied based on the subject's individual capacity to discriminate asynchronous from synchronous stimuli. We defined these three difficulty levels as the SOAs at which the subjects easily detected asynchrony (easy-difficulty), were able to detect asynchrony at approximately chance (medium-difficulty), and rarely were able to detect asynchrony (hard-difficulty). To vary the task difficulty, we individually determined the training SOAs based on the participants' performance for the pre-training assessment. We employed fitting method that has been applied to derive estimates of temporal acuity in numerous prior studies (Noel, Łukowska, Wallace, & Serino, 2016; Powers et al., 2009; R. A. Stevenson & Wallace, 2013). Individual participants' data on pre-training SJ assessment was fit with two sigmoid curves using the *glmfit* function in MATLAB to separately fit the data for the auditory-leading stimuli (utilized to estimate the auditory-leading TBW) and visual-leading stimuli (utilized to estimate visual-leading TBW). Using the estimates of the fitted data, the training SOAs were defined by the SOA (i.e. x value)

at which the best-fit sigmoids  $y$  value equaled a value of 10, 20, and 30% report of synchrony for the easy-difficulty group; 40, 50, and 60% report of synchrony for the medium-difficulty group; and 70, 80, and 90% report of synchrony for the hard-difficulty group (Figure 3-1C). Thus, we determined six asynchronous training SOAs (3 auditory-leading; 3 visual-leading) from the individuals' perceptual report while also presenting the truly synchronous condition.

The number of trials per SOA was not equally distributed during the training sessions. The veridical simultaneous condition had a 6:1 ratio to the asynchronous conditions such that the total number of simultaneous trials presented was equal to the total number of asynchronous trials presented. We presented an equal number of synchronous and asynchronous trials to mitigate concerns about creating a response bias.

#### *Grand Mean SOA analysis*

We first characterized the effect of perceptual training on the perceptual report of synchrony in the SJ task by measuring how the overall group probability of report of synchrony changed as a function of training status and SOA. We measured individual means at each SOA that were then averaged to produce grand averages. A repeated-measures ANOVA was conducted for each group to determine if a significant training status x SOA interaction existed. An analysis of simple effects was conducted if a significant training status x SOA interaction existed. *Post hoc* pairwise comparisons of simple effects were conducted with Bonferroni correction to determine if the training session for each group significantly altered the report of synchrony at each SOA. Degrees of freedom were adjusted for all interactions where sphericity was violated using a Greenhouse-Geisser correction.

#### *TBW Analysis*

Estimates of the whole, auditory-leading, and visual-leading TBWs were calculated for pre- and post-training SJ assessments using the same fitting method as described for

determining the individualized training SOAs. A criterion at which the best-fit sigmoids'  $y$  value equaled 75% report of synchrony was utilized to estimate the TBW. The auditory-leading TBW was defined as epoch for which the auditory-leading sigmoid fit's  $y$  value equaled 75% report of synchrony to the point of subjective simultaneity (PSS, i.e. the SOA at which the participant is most likely to report the stimuli as synchronous), the point at which the sigmoid fits crossed; the visual-leading TBW was defined as the epoch for which the visual-leading sigmoid fit's  $y$  value equaled 75% report of synchrony to the PSS. The whole TBW was defined as the sum of the auditory-leading and visual-leading TBWs. Group-level analysis of the TBW was taken by calculating the mean auditory-leading, visual-leading, and whole TBW individually fit for each participant. To determine within each group if there was an effect of perceptual training, the pre-training and post-training TBWs were compared using a paired samples  $t$ -test, consistent with prior analyses of perceptual training. In order to compare the training effect across each group, we first determined the mean change in the TBW. We then conducted a one-way between subjects ANOVA to assess whether the change in the TBW differed across training difficulty levels.

## Results

### *Perceptual Training with Easy Training Stimuli Increases Report of Simultaneity while Difficult Training Stimuli Decrease Report of Simultaneity*

To determine if perceptual training altered simultaneity judgments differentially dependent upon the difficulty of the training, we conducted separate SOA x pre-/post-training status within-subjects repeated measures ANOVAs for each group (i.e. easy-, medium-, and hard-difficulty training protocols). Thus, the measure of interest was whether a significant SOA x training status interaction was present. A significant interaction was observed for the easy-difficulty group ( $F_{6.750, 107.993} = 3.692, p = .001$ ) and hard-difficulty group ( $F_{5.832, 93.318} = 2.507, p = .028$ ). A significant interaction was not observed for the medium-difficulty group ( $F_{4.706, 75.298} =$

1.175,  $p = .330$ ). *Post hoc* analysis of simple main effects was conducted with Bonferroni correction for the simple effect of training status at each SOA. For the easy-difficulty group (Figure 3-2A), an analysis of simple main effects showed that perceptual training increased the probability of reporting simultaneity at SOAs of -400 ms ( $p = .017$ ) and -300 ms ( $p = .024$ ) for which the auditory lead was the greatest. In contrast, for the hard-difficulty group (Figure 3-2C), perceptual training decreased the probability of reporting simultaneity at SOAs of -150 ms ( $p = .012$ ), -100 ms ( $p < .001$ ), 50 ms ( $p = .003$ ), 100ms ( $p = .001$ ), 150 ms ( $p < .001$ ), 200 ms ( $p = .049$ ), 250 ms ( $p = .014$ ), and 300 ms ( $p = .026$ ). Thus, these results illustrate that training difficulty had opposing effects on audiovisual simultaneity judgments, and suggest that training on more difficult judgments may improve temporal acuity and narrow the TBW.

#### *Effect of Training Difficulty on TBW*

We next sought to determine the effect of each training protocol on the TBW within each group by conducting paired samples *t*-tests for the auditory-leading, visual-leading, and whole (i.e., left and right sides) TBWs (Figure 3-3). Analysis of the whole TBW revealed that only for those individuals in the hard-difficulty group did narrowing following perceptual training (from 440 ms to 329 ms;  $t(16) = 2.669$ ,  $p = .017$ ). In contrast, those in the medium-difficulty (from 386 ms to 393 ms;  $t(16) = -.165$ ,  $p = .871$ ) and easy-difficulty (from 461 ms to 569 ms;  $t(16) = -2.044$ ,  $p = .058$ ) groups were not significantly affected by perceptual training (although the easy-difficulty group approached a significant increase in the TBW).

We next determined if the auditory-leading and visual-leading TBWs were differentially altered by perceptual training difficulty. For the easy-difficulty group, the auditory-leading TBW significantly increased following perceptual training (from 178 ms to 258 ms;  $t(16) = -2.553$ ,  $p = .021$ ) while the visual-leading TBW was not altered (from 283 ms to 310 ms;  $t(16) = -.722$ ,  $p = .481$ ). For the medium-difficulty group, neither the auditory-leading TBW (from 171 to 161;  $t(16) = 1.249$ ,  $p = .230$ ) nor the visual-leading TBW (from 214 ms to 232 ms;  $t(16) = -.451$ ,  $p = .658$ )

were altered by perceptual training. For the hard-difficulty group, the auditory-leading TBW significantly decreased following perceptual training (from 188 ms to 153 ms;  $t(16) = 2.991$ ,  $p = .009$ ) while the visual-leading TBW was not altered (from 212 ms to 176 ms;  $t(16) = 1.797$ ,  $p = .091$ ).

### *Greater Enhancement of Temporal Acuity with More Difficult Training Protocols*

Although we previously determined if each group experienced a change in temporal acuity following training (i.e the  $\Delta$ TBW was significantly different from 0) by individual paired samples  $t$ -tests, we sought to compare if the relative change in the TBW differed when compared across groups. To determine if the change in the TBW ( $\Delta$ TBW = post-training - pre-training) differed by difficulty of the training protocol (Figure 3-3D) we conducted a one-way between-subjects ANOVA. This analysis demonstrated a significant effect of training protocol difficulty ( $F_{2,48} = 4.498$ ,  $p = .016$ ). *Post hoc* comparisons with Bonferroni correction revealed that  $\Delta$ TBW for the easy-difficulty group (107 ms) was significantly different from the hard-difficulty group (-71 ms), ( $p = .013$ ). Thus, the decrease in the TBW for individuals completing the hard-difficulty protocol resulted in relatively greater enhancement of temporal acuity in comparison to both within-group pre-training levels as well as relative to those individuals completing perceptual training with the easy difficulty protocol. In contrast, the  $\Delta$ TBW for the medium-difficulty group (7 ms) did not differ relative to either the easy-difficulty group ( $p = .298$ ) or hard-difficulty group ( $p = .587$ ).

## **Discussion**

In the current study, we observe that task difficulty during perceptual training is a critical element for eliciting multisensory perceptual learning. We find that only the perceptual training protocol that employed asynchronous audiovisual stimuli for which individuals had the greatest difficulty detecting asynchrony resulted in an enhancement of temporal acuity. Surprisingly, we

also find that the training protocol for which participants were able to most easily detect the asynchrony actually impaired temporal acuity following perceptual training (i.e., resulted in an enlargement in the TBW). Our results suggest that perceptual training protocols to enhance multisensory temporal acuity can be optimized by employing only stimuli during training for which it is difficult for individuals to discriminate synchronous and asynchronous stimuli.

Numerous factors, including stimulus properties and reinforcement signals, interact to promote perceptual learning (Aaron R Seitz & Dinse, 2007). While passive exposure to asynchronous stimuli has been previously observed to alter the perceptual representation of perceived synchrony, both over the course of many trials (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Van der Burg, Alais, & Cass, 2015; Vroomen, Keetels, De Gelder, & Bertelson, 2004) and trial-to-trial (Van der Burg, Alais, & Cass, 2013; Van der Burg, Orchard-Mills, & Alais, 2015), passive exposure to asynchronous does not seem to enhance temporal acuity. Although these prior studies, along with those studies that observe enhancement of temporal acuity with feedback (Powers et al., 2009), suggest that a reinforcement signal was needed to enhance multisensory temporal acuity, the properties of the stimuli employed during perceptual training have not been systematically explored (Powers et al., 2009). Our findings suggest that while feedback may be an important element for enhancing multisensory temporal acuity, there is an interaction between the feedback signal and difficulty of the perceptual training task. *Powers et al.* (2009) first demonstrated that perceptual training enhances temporal acuity by using a fixed set of SOAs for asynchronous stimuli during training sessions that were independent of an individual's temporal acuity. In this study, they reported that perceptual training did not enhance temporal acuity for a subset of the participants whose TBWs tended to be narrower prior to perceptual training relative to other participants for which training enhanced temporal acuity. Based on the results of the current study, we hypothesize that the SOAs employed by *Powers et al.* during perceptual training sessions were of insufficient difficulty to elicit perceptual learning in this subset of participants who failed to demonstrate training-mediated changes. This, along

with our findings, suggests that individual differences in temporal acuity prior to perceptual training must be accounted for to optimize the effects of perceptual learning.

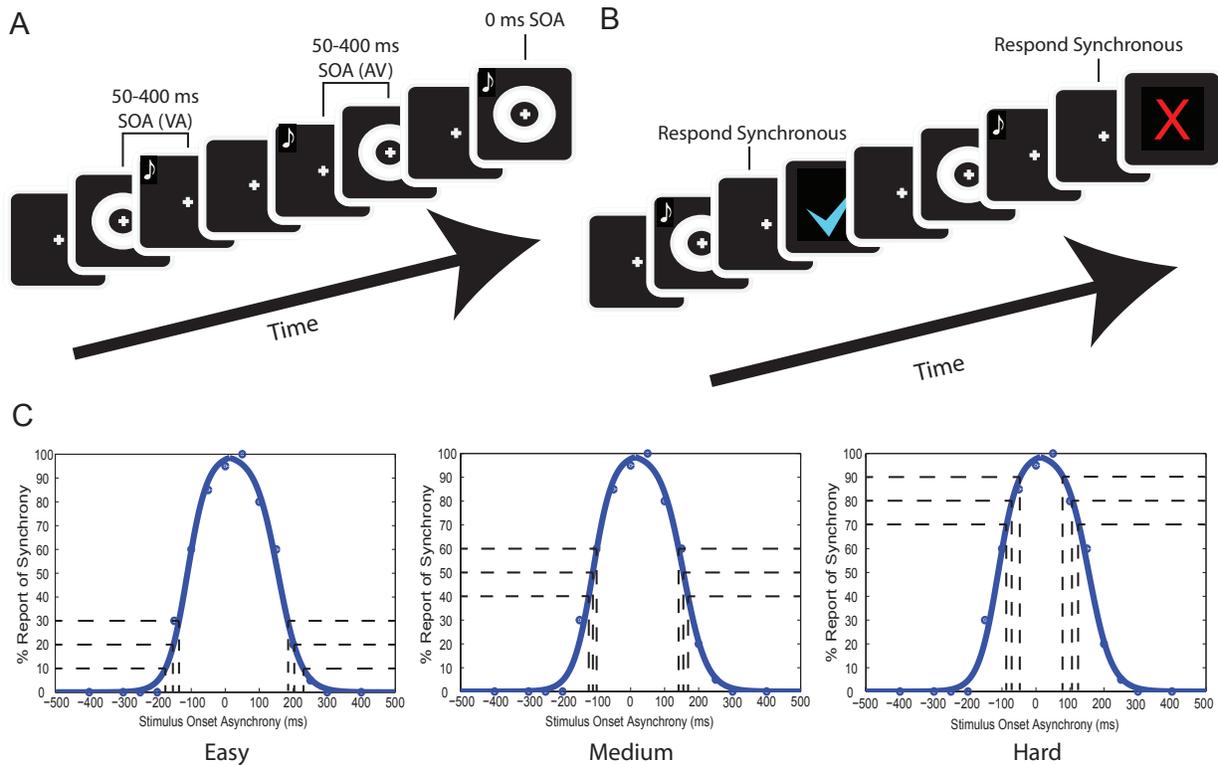
We were surprised to observe that for those in the easy-difficulty protocol, presenting SOAs during training for which asynchrony was easily detected increased the report of synchrony at the SOAs with the greatest auditory-lead, thus resulting in a widening of the auditory-leading TBW. While this finding initially seemed surprising as we expected that the presence of a feedback signal during the training session would either maintain (i.e. not alter) or enhance temporal acuity, it supports the earlier report by Powers *et al.* (2009) who observed that some individuals with narrow TBWs prior to training experienced an increase in their TBW following training in their protocol (Powers *et al.*, 2009). One possible explanation for this paradoxical effect is that the feedback this group received was generally not informative in shaping their perceptual decisions, but rather confirmed these decisions. The lack of an informative feedback signal may be similar to the task without feedback, making these individuals more susceptible to adaptation effects. Prior studies shown that extensive, passive exposure (i.e., in the absence of feedback) to asynchronous stimuli may widen the TBW (Fujisaki *et al.*, 2004; Navarra, Soto-Faraco, & Spence, 2007; Navarra *et al.*, 2005; Vroomen *et al.*, 2004). Furthermore, Powers *et al.* (2009) report an increase in the TBW for individuals who were passively exposed to the same stimuli as those who received feedback during the training session (Powers *et al.*, 2009). An alternative explanation is that the shift in the auditory-leading TBW for the easy-difficulty group resulted from an adaptation effect produced by presenting particularly large SOAs. Interestingly, passive exposure to audiovisual stimuli with auditory-leading SOAs is observed to tend to elicit the greatest change in temporal representations (Fujisaki *et al.*, 2004; Vroomen *et al.*, 2004). It is possible that relative to the other conditions, the large SOAs at which participants could easily detect asynchrony were sufficient enough to widen the TBW. Although the distribution appears to be shifted towards an auditory-leading stimulus for this portion of the TBW, the majority of the effect is observed at only the most

auditory SOAs suggesting the change we observe is not a true adaptation effect as the auditory-leading SOAs are not uniformly affected following perceptual training.

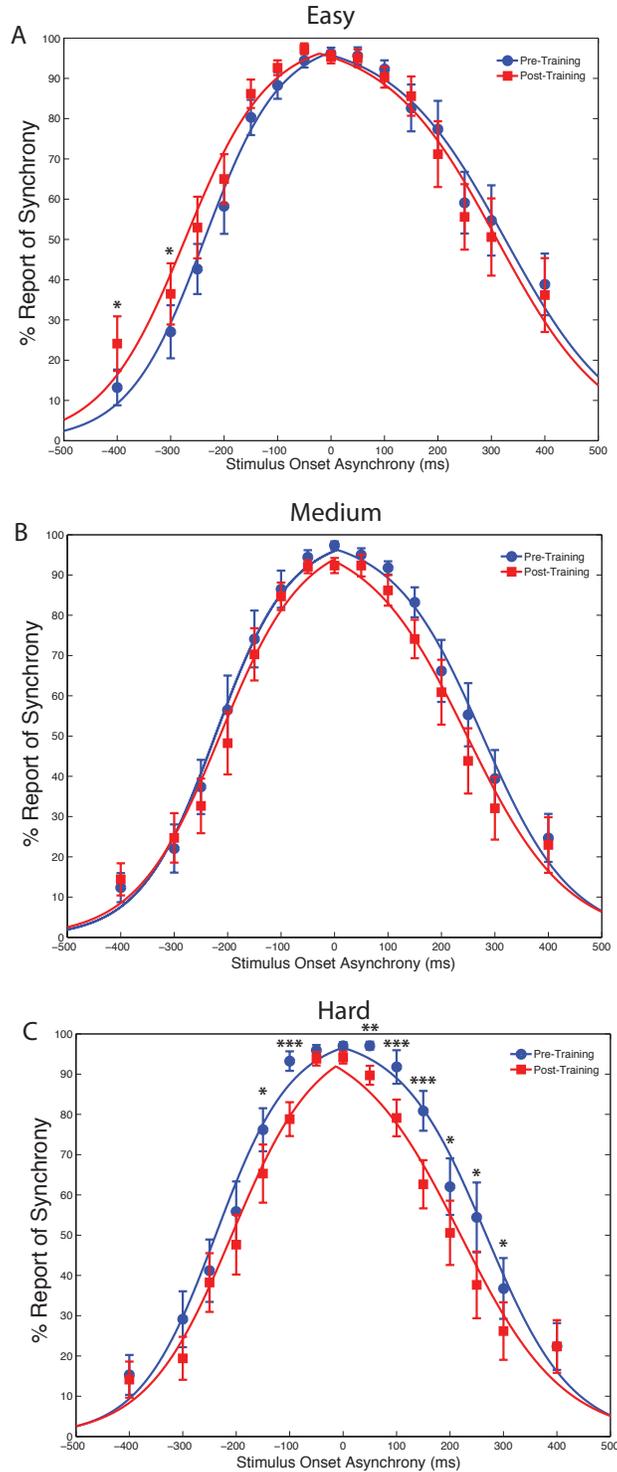
Our findings collectively suggest that in order to optimize perceptual training protocols for enhancing temporal acuity, the difficulty of the perceptual training task as well as pre-training differences in temporal acuity must be considered. Based on our findings, one area of future research to optimize perceptual training would be to determine if increasing the task difficulty following an initial improvement would further enhance temporal acuity. Progressively adaptive training methods might be the ideal method for improving temporal acuity. Another issue for future research to address is possibility that increasing task difficulty impairs generalization of the changes in temporal acuity to stimuli with different properties as has been observed for visual perceptual learning. Although the reverse hierarchy theory of visual perceptual learning would suggest that if multisensory perceptual learning occurs at lower-levels of sensory processing, difficult tasks would not generalize (Ahissar & Hochstein, 2004), evidence suggests that multisensory perceptual learning engages higher order cortical regions, such as the posterior superior temporal sulcus (pSTS), and thus may not exhibit entirely similar properties as unisensory perceptual learning (Powers et al., 2012). Determining how to improve perceptual training by understanding these additional contributions of task difficulty will be important in tailoring future training regimens, particularly in efforts to use such training to improve perceptual function in those with disorders such as autism and dyslexia.

### **Acknowledgements**

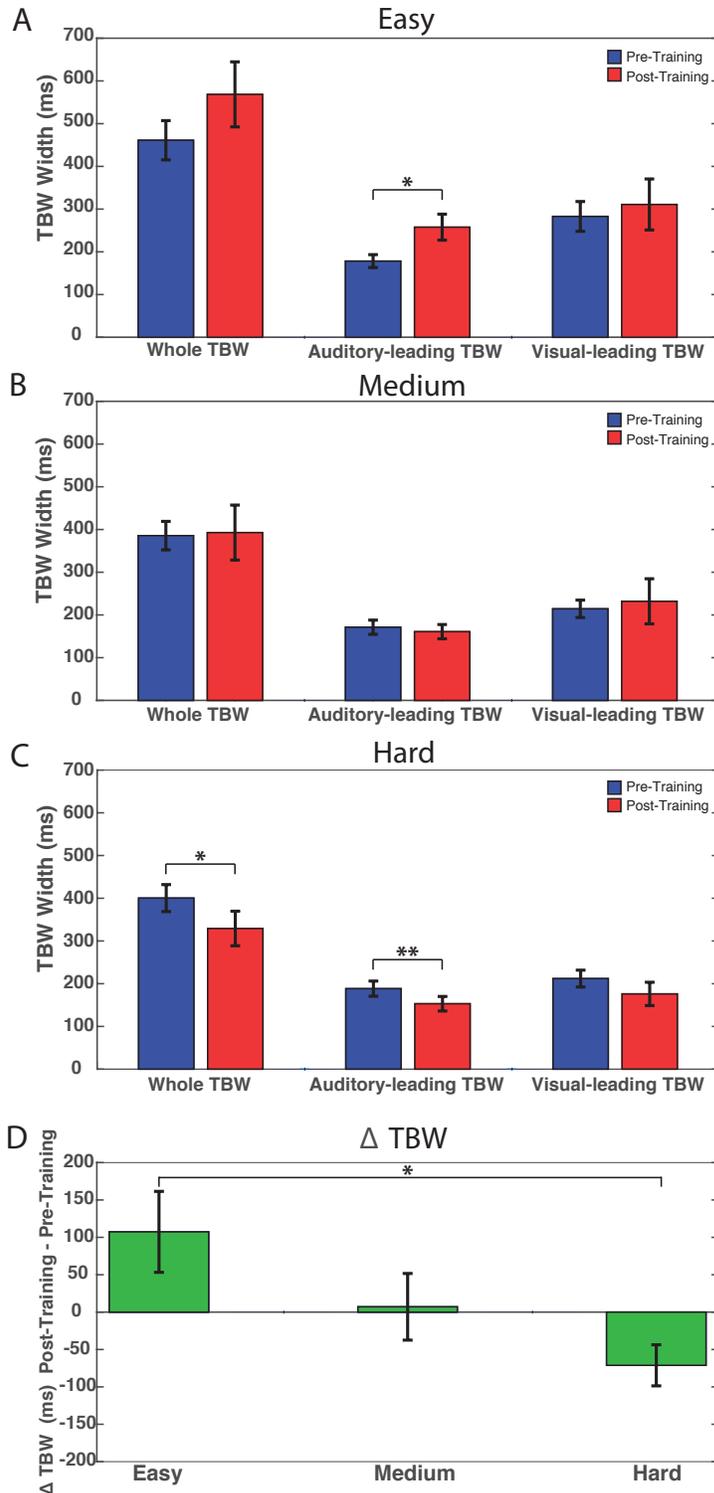
The project was supported by NIH grants CA183492 and HD083211, the Simons Foundation Autism Research Initiative and the Wallace Foundation. The authors have no conflicts of interest to declare.



**Figure 3-1. SJ Task and Study Methods** a) A simultaneity judgment (SJ) task was utilized to measure the TBW prior to and following perceptual training. b) Illustration of correct feedback given for a response of synchronous after an objectively synchronous trial followed by subsequent asynchronous trial for which incorrect feedback is presented for a response of synchronous. c) Perceptual training sessions were individualized by determining training SOAs at three different levels of difficulty. Pre-training SJ assessment data from a representative participant (circles, report of synchrony at each SOA; solid lines, individual sigmoid curve fit to data) illustrates how training SOAs would have been selected if the participant was randomly assigned to the either the easy- (left), medium- (center), or hard- (right) difficulty condition. The value of the asynchronous training SOAs corresponds to the point at which the vertical dashed lines intersect the x-axis.



**Figure 3-2. Effect of Task Difficulty on the Report of Synchrony at each SOA** a) An increase in the report of synchrony at the most auditory-leading SOAs was reported for the easy-difficulty group. b) No change in the report of synchrony was observed for the medium difficulty group. c) A decrease in the report of synchrony at both auditory-leading and visual-leading SOAs was observed. Circles indicate pre-training and squares indicate post-training mean report of synchrony. Error bars, SEM. (\*,  $p < .05$ , \*\*,  $p < .01$ , \*\*\*,  $p < .001$ ).



**Figure 3-3. Change in TBW for Each Group** a) For the easy-difficulty group, an increase in the auditory-leading TBW was observed post-training. b) No change in the TBW was observed for the medium-difficulty group. c) A decrease in the whole and auditory-leading TBW was observed post-training for the hard-difficulty group. d) A comparison of the  $\Delta$ TBW between groups determined that relative  $\Delta$ TBW differed between easy-difficulty and hard-difficulty groups. In contrast the  $\Delta$ TBW for the medium-difficulty group was not significantly different from either the easy-difficulty or hard-difficulty groups. Error bars, SEM. (\*,  $p < .05$ , \*\*,  $p < .01$ ).

## References

- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631), 401-406.
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences*, 8(10), 457-464.
- Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., . . . Johnston, E. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501(7465), 97-101.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433-436.
- Chen, Y.-C., Shore, D. I., Lewis, T. L., & Maurer, D. (2016). The development of the perception of audiovisual simultaneity. *Journal of experimental child psychology*, 146, 17-33.
- DeLoss, D. J., Watanabe, T., & Andersen, G. J. (2014). Optimization of perceptual learning: Effects of task difficulty and external noise in older adults. *Vision research*, 99, 37-45.
- Deveau, J., Ozer, D. J., & Seitz, A. R. (2014). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology*, 24(4), R146-R147.
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, 66(8), 1388-1404.
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9(6), 719-721.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature neuroscience*, 7(7), 773-778.
- Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, 49(3), 461-467.  
doi:10.1016/j.neuropsychologia.2010.11.041

- Hillock - Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5), 688-696.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, *7*(10), 3215-3229.
- Mishra, J., Anguera, J. A., & Gazzaley, A. (2016). Video Games for Neuro-Cognitive Optimization. *Neuron*, *90*(2), 214-218.
- Mishra, J., de Villers-Sidani, E., Merzenich, M., & Gazzaley, A. (2014). Adaptive training diminishes distractibility in aging across species. *Neuron*, *84*(5), 1091-1103.
- Mishra, J., & Gazzaley, A. (2014). Harnessing the neuroplastic potential of the human brain & the future of cognitive rehabilitation. *Frontiers in human neuroscience*, *8*, 218.
- Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience letters*, *413*(1), 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. *Cognitive brain research*, *25*(2), 499-507.
- Noel, J.-P., Łukowska, M., Wallace, M., & Serino, A. (2016). Multisensory simultaneity judgment and proximity to the body. *Journal of Vision*, *16*(3), 21-21.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*(4), 437-442.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, *32*(18), 6263-6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, *29*(39), 12265-12274.

- Powers III, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports*, 6, 23374.
- Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current opinion in neurobiology*, 17(2), 148-153.
- Seitz, A. R., Nanez, J. E., Sr., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, 6(9), 966-973. doi:10.1167/6.9.9
- Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia*, 61, 259-268.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4), 255-266.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15(8), 520-535.
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*, 34(3), 691-697.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, 227(2), 249-261. doi:10.1007/s00221-013-3507-3
- Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research*, 225(4), 479-489. doi:10.1007/s00221-012-3387-y

- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. *Journal of Experimental Psychology-Human Perception and Performance*, *38*(6), 1517-1529. doi:10.1037/a0027339
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26*(2), 212-215.
- Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid Recalibration to Audiovisual Asynchrony. *Journal of Neuroscience*, *33*(37), 14633-14637. doi:10.1523/jneurosci.1182-13.2013
- Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific reports*, *5*.
- Van der Burg, E., Orchard-Mills, E., & Alais, D. (2015). Rapid temporal recalibration is unique to audiovisual stimuli. *Experimental Brain Research*, *233*(1), 53-59.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871-884.
- Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive brain research*, *22*(1), 32-35.
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105-123.
- Wang, Y., Song, Y., Qu, Z., & Ding, Y. (2010). Task difficulty modulates electrophysiological correlates of perceptual learning. *International Journal of Psychophysiology*, *75*(3), 234-240.

## Chapter IV

### The Impact of Feedback on the Different Time Courses of Multisensory Temporal Recalibration

#### Abstract

The capacity to rapidly adjust perceptual representations confers a fundamental advantage when confronted with a constantly changing world. Unexplored is how feedback regarding sensory judgments (top-down factors) interacts with sensory statistics (bottom-up factors) to drive long- and short-term recalibration of multisensory perceptual representations. Here, we examined the time course of both cumulative and rapid temporal perceptual recalibration for individuals completing an audiovisual simultaneity judgment task in which they were provided with varying degrees of feedback. We find that in the presence of feedback (as opposed to simple sensory exposure) temporal recalibration is more robust. Additionally, differential time courses are seen for cumulative and rapid recalibration dependent upon the nature of the feedback provided. Whereas cumulative recalibration effects relied more heavily on feedback that informs (i.e., negative feedback) rather than confirms (i.e., positive feedback) the judgment, rapid recalibration shows the opposite tendency. Furthermore, differential effects on rapid and cumulative recalibration were seen when the reliability of feedback was altered. Collectively, our findings illustrate that feedback signals promote and sustain audiovisual recalibration over the course of cumulative learning and enhance rapid trial-to-trial learning. Furthermore, given the differential effects seen for cumulative and rapid recalibration, these processes may function via distinct mechanisms.

## Background

In order to accurately perceive the world, individuals must adjust their perceptual representations to meet the changing nature of the sensory world and changing task contingencies (Ghose, 2004). Given this, the capacity to rapidly adjust perceptual representations confers a fundamental advantage (Medina & Rapp, 2014; Yu, Stein, & Rowland, 2009). Such perceptual plasticity often leads to an improved representation of the sensory environment, a process termed perceptual learning (Manfred Fahle, 2005; Shams & Seitz, 2008). Changes in perceptual representations resulting from perceptual learning have been observed to occur within both rapid (M Fahle, Edelman, & Poggio, 1995; Wozny & Shams, 2011) and more gradual time courses (Manfred Fahle & Edelman, 1993; Watanabe et al., 2002). Furthermore, the contribution of feedback signals, in conjunction with sensory experience, is known to alter the rate of perceptual learning or enable perceptual learning to occur when sensory experience is insufficient (Aaron R Seitz & Dinse, 2007).

Although initial investigations of perceptual plasticity tended to focus on changes in perception for a single sensory modality, there has been an increasing interest in examining the plasticity of multisensory perceptual representations (Shams & Seitz, 2008; Wallace & Stevenson, 2014). One such area of investigation has focused on how the temporal processing of multisensory stimuli (particularly audiovisual stimuli (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Keetels & Vroomen, 2007; Van der Burg, Alais, & Cass, 2013; Vroomen, Keetels, De Gelder, & Bertelson, 2004); for other modalities see (Keetels & Vroomen, 2008; Navarra, Soto-Faraco, & Spence, 2007)) can be altered via changes in sensory experience. The temporal structure of sensory stimuli from the different modalities is a fundamental feature determining whether these stimuli should be associated or perceived as a single multisensory event (Noel, Łukowska, Wallace, & Serino, 2016; Stein, Stanford, & Rowland, 2014; Vroomen & Keetels, 2010). One critical aspect of this process must take into account the differences in neural and physical transmission times for the respective sensory stimuli (e.g., light and sound energy

propagate through the environment at very different rates). In order to circumvent this challenge and ultimately achieve perceptual coherence, there exists an epoch of time spanning several hundred milliseconds within which stimuli from vision and audition are likely to be associated. This construct has been collectively referred to as the temporal binding window (TBW). Similarly, the point of asynchrony at which the separate sensory stimuli are most likely to be perceived as occurring synchronously has been termed the point of subjective simultaneity (PSS). These two metrics, the TBW and PSS, are thus important tools in evaluating the nature of audiovisual temporal representations.

Prior work has shown both the TBW and PSS to be malleable. These dynamic changes in the TBW and PSS, termed temporal recalibration, were initially hypothesized as a means to resolve asynchronous sensory signals reflective of the statistics of the environment (Vroomen & Keetels, 2010). Thus, initial studies showed that it was possible to shift an individual's PSS by providing extensive experience that overrepresented certain asynchronies (Fujisaki et al., 2004; Vroomen et al., 2004). More recent evidence suggests that these changes not only occur after extensive experience, but can also be seen on a moment-to-moment basis (i.e., based on the characteristics of the previous trial,  $t-1$ ) (Bruns & Röder, 2015; Harvey, Van der Burg, & Alais, 2014; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015; Wozny & Shams, 2011). Thus, changes in multisensory temporal representations happen on both rapid and cumulative time scales. Such observations raise fundamental mechanistic questions about these short- and longer-term changes, most immediately in regards to whether one (short-term) represents the substrate upon which the other (longer-term) is built. One can envision a scenario in which rapid temporal recalibration may be needed in order to properly represent immediate changes in the sensory environment whereas cumulative temporal recalibration may result in more durable changes in perceptual representations (Bruns & Röder, 2015; Van der Burg, Alais, & Cass, 2015).

While sensory experience is undoubtedly an important element that influences perceptual plasticity, feedback signals that inform an individual regarding the accuracy of their perceptual judgments are likely to interact with sensory experience to influence temporal recalibration of the TBW and PSS. Early studies of visual perceptual learning suggest that feedback signals enhance perceptual learning (Manfred Fahle & Edelman, 1993; Aaron R. Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006) and are capable of eliciting perceptual learning even in the absence of awareness in regards to the changing nature of the sensory environment (Aaron R Seitz, Kim, & Watanabe, 2009). Increasingly, changes on top-down processing regions have been observed to parallel perceptual learning, (Law & Gold, 2008; Law & Gold, 2009) and are likely to be activated by a feedback signal. Collectively, the evidence suggest that the dynamics of perceptual learning are likely dependent upon coordinated interactions between sensory statistics primarily represented in low-level cortical areas and the brain areas that initially represent them and higher-order factors and their neural substrates (Watanabe & Sasaki, 2015). Recent studies have observed that feedback signals also produce rapid improvements in multisensory temporal acuity (Powers, Hillock, & Wallace, 2009; Powers III, Hillock-Dunn, & Wallace, 2016) and elicit changes in connectivity between primary sensory cortices and multisensory cortex (Powers, Hevey, & Wallace, 2012). Despite the independent evidence for the importance of these bottom-up and top-down factors in perceptual plasticity, few studies have looked at the interdependence between them. Here, we sought the interaction of bottom-up and top-down factors in perceptual plasticity by altering top-down factors (i.e., presence of a feedback signal, feedback reliability) and examining its impact on temporal recalibration across both immediate and longer-term time scales.

## **Methods**

### *Participants*

Sixty-five young adults partook in this study (36 females; Age,  $M = 20.48$  years, Range = 18 – 28 years). All participants had self-reported normal hearing and normal or corrected to normal vision. Written informed consent was obtained from all individuals participating in this study. All participant recruitment and experimental procedures were approved by the Vanderbilt University Institutional Review Board and were in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### *Assessment of Temporal Acuity by Simultaneity Judgment Task*

We employed simultaneity judgment (SJ) task to measure audiovisual temporal acuity as prior studies assessing temporal recalibration have utilized similar SJ tasks (Fujisaki et al., 2004; Harvey et al., 2014; Stevenson & Wallace, 2013; Van der Burg et al., 2013). Participants were seated in a light and sound attenuating WhisperRoom™ (SE 2000 Series, Whisper Room Inc.) room for all tasks. All visual stimuli were presented at approximately 60 cm from the seated participants. A fixation marker (1cm x 1cm) on a black background was present on the screen both between trials and throughout the duration of a trial including presentation of the visual stimulus. Participants were asked to maintain fixation on the fixation marker throughout the experiment. For the SJ task, participants were instructed to judge whether the visual stimulus and auditory stimulus “were synchronous, at the same time” or “were asynchronous, at different times” by pressing either 1 or 2 respectively using a keyboard (see Figure 4-1A). The visual stimulus consisted of a white ring on a black background that subtended  $7.2^\circ$  of visual space with an outer diameter of 6.0 cm and an inner diameter of 3.0 cm. Visual stimuli were presented for 8.3 ms (the duration of a single screen refresh cycle) on a monitor (Samsung syncmaster 22 inch 2233 RZ LCD) with a refresh-rate of 120 Hz. The auditory stimulus consisted of an 1800 Hz tone that was presented binaurally via headphones (Sennheiser HD 558) with no interaural

time or level differences. Auditory stimuli were presented at 83 dB and were calibrated using a sound level meter (Larson Davis SoundTrack® LxT2). For each trial, visual and auditory stimuli were presented in synchrony (0 ms of asynchrony) or with a stimulus onset asynchrony (SOAs) ranging from  $\pm 400$ -50 ms (negative values indicate that the auditory stimulus was the leading stimulus while positive values indicate that visual stimulus was the leading stimulus). To ensure accurate presentation of auditory and visual stimuli, SOAs were verified externally using an oscilloscope. A response screen was presented following presentation of each audiovisual pair at which time subjects could make a response. The inter-trial interval (ITI) between trials was randomly jittered from 500 to 1500 ms (uniform distribution). MATLAB (The MathWorks, Inc.) with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) was used to create and present the SJ task.

#### *General Experimental Procedure for the Assessment of Temporal Recalibration*

To derive initial estimates of the TBW and PSS (to assess cumulative recalibration) and trial-to-trial change in the PSS and TBW (to assess rapid recalibration), participants first completed an initial block of 300 trials of SJ task comprised of 20 trials at each of the following SOAs:  $\pm 400$ , 300, 250, 200, 150, 100, 50, and 0 ms. Performance over this block of trial was utilized to derive Trial 1 of the time course analysis. Participants subsequently completed a second block of a total 720 trials of the SJ task with stimuli presented at SOAs of  $\pm 150$ , 100, 50, and 0 ms. To avoid introducing a response bias for participants in groups receiving feedback, the number of trials at each SOA was not equally distributed in the second trial block. Instead true synchrony was presented at a 6:1 ratio to the objectively asynchronous SOAs such that the total number of simultaneous trials presented was equal to the total number of asynchronous trials presented (0 ms, 360 trials;  $\pm 150$ , 100, and 50 ms x 60 trials each). Participants next completed a third block of 300 additional trials of the SJ task identical to the first block of 300 trials (20 trials x 15 SOAs;  $\pm 400$ , 300, 250, 200, 150, 100, 50, and 0 ms). The time course

analysis was performed using the responses in the second and third blocks. During only the second block was feedback presented for groups receiving a form of feedback. No feedback was presented during the first or third blocks. Participants were given an opportunity to rest after every 100 trials in each experimental trial block.

### *Presentation of Feedback Signal*

For all participants, feedback was not provided during the presentation of trials in the first trial block. Participants were randomly assigned to one of four experimental groups characterized by the nature of the feedback presented during the second block. For the first group ( $n = 15$ ), participants did not have access to any explicit feedback. Participants in the second group ( $n = 25$ ) had access to reliable visual feedback in the form of a blue-green check mark or red X following objectively correct and incorrect responses. Reliable feedback was defined as feedback that accurately reflected the objective relationship of the audiovisual stimuli. The third ( $n = 13$ ) and fourth ( $n = 12$ ) groups were presented with reliable feedback on only 80% and 50% of trials (i.e. false or erroneous feedback on 20% and 50% of the remaining trials). False feedback (i.e. feedback that was not reliable) was defined as the presentation of the incorrect feedback for each SOA-response pair (i.e. a response of synchronous for a trial in which the SOA presented was 0 ms yielded the presentation of a red X, the exact opposite of the objectively accurate feedback). All feedback was presented for 500 ms immediately following the participant's response. No feedback was presented for the third trial block for all participants (post-feedback period as denoted by dashed line at Trial 721 in Figures 4-2, 4-3, 4-4, 4-5, and 4-6).

### *Analysis of the Time Course of Multisensory Temporal Recalibration*

Two distinct multisensory temporal recalibration time courses are of interest here. A first, denominated 'cumulative' adaptation, refers to the degree to which participants consider

accumulating feedback when executing audio-visual simultaneity judgments. This adaptation, thus, requires the conscious acknowledgment of received feedback. The second, referred to as ‘rapid’ recalibration, denotes to the degree to which the nature of the immediately preceding trial (t-1) – audio- or visual-leading – influences the perception of simultaneity at the given trial (t). The examination of rapid audio-visual temporal recalibration effects, thus, is taken to index an implicit sensory phenomenon, perceptual learning, and involves a one-back analysis (analysis of trial t as a conditional of trial t-1).

To maintain a consistent estimate of the TBW and PSS across all trial blocks, only SOAs of  $\pm 150$ , 100, 50, and 0 ms were utilized to fit distributions for the entire time course. Although we employed wider ranging SOAs in the first trial block to ensure an accurate estimate of the PSS and TBW, further analysis revealed that fitting these distributions with the entire course of SOAs and those present across all trial blocks did not result in any significant differences in our initial estimates for all measures ( $p > .05$  for all measures). Thus, initial distributions (Trial 0 in the time course analysis) were drawn based on the 140 trials in the first trial block. These distributions were fitted with a Gaussian normal distribution whose amplitude, mean, and standard deviation were free to vary (See Eq. 1). The normal distribution proved to be a good fit (mean  $R^2 = 85.6$ , SD = 2.45). The mean of the best fitting distribution is taken as the PSS and the standard deviation as a measure of the TBW. That is, PSS is the point (i.e., SOA) at which participants are most likely to categorize a presentation as synchronous and the TBW is the temporal interval over which participants are highly likely to categorize the presentation as synchronous.

$$P(\text{response}|\text{SOA}) = \text{amp} X \exp^{-\left(\frac{(\text{SOA}-\text{PSS})^2}{2\text{SD}^2}\right)} \quad (\text{Eq. 1})$$

In order to index rapid recalibration, the amount of change in these values (PPS and TBW) is computed (e.g.,  $\Delta\text{PPS} = \text{PPS audio-leading} - \text{PSS visual leading}$ ) as a function of the prior trial (either correct or incorrect). Further, in order to scrutinize the time courses of rapid vs. slow multisensory temporal recalibration effects, we adopt a sliding-window approach (Figure 1B). That is, after the first estimation of the mean and standard deviation of the Gaussian describing reports of synchrony, a window of 140 trials – initially placed between Trials 1 and 140 – is moved trial-per-trial across the span of the second trial block (Trials 1-720) as well as during the third trial block post-feedback (Trials 721-860). At each step the new distribution is fitted again, and estimates of the mean and standard deviation are calculated. Similarly, at each step rapid recalibration values are computed. Upon completion of the protocol (~ 5 000 fittings per subject), PSS and TBW values were normalized from 0 to 1 within-subjects.

To assess if the time course of temporal recalibration was effected by positive or negative feedback, distributions of perceived simultaneity (i.e., report of synchrony) as a function of SOA were compiled for each participant separately for the cases in which on the precedent trial (t-1) participants were informed that their answer had been correct (t-1 correct; prior positive feedback) or incorrect (t-1 incorrect; prior negative feedback). Additionally, in order to compute rapid recalibration effects, reports of synchrony were further bifurcated into those in which trial t-1 had either a negative (i.e., audition led) or positive SOA (i.e., vision led).

## **Results and Discussion**

### *Feedback Accelerates and Maintains Cumulative Temporal Recalibration*

Previous adaptation studies have shown that temporal recalibration occurs over slow time scales as extended periods of passive exposure to asynchronous stimuli (often biased in the direction of either an auditory or visual leading stimulus with a constant SOA) elicits changes in perceptual representations as indexed via the PSS and TBW (audiovisual see (Bruns & Röder, 2015; Fujisaki et al., 2004; Harvey et al., 2014; Noel, Wallace, Orchard-Mills, Alais, &

Van der Burg, 2015; Van der Burg et al., 2013; Van der Burg, Alais, et al., 2015; Van der Burg & Goodbourn, 2015; Van der Burg, Orchard-Mills, & Alais, 2015; Vroomen et al., 2004), other modalities see (Van der Burg et al., 2013)). Here, we sought to address if sensory experience for unbiased, asynchronous stimuli elicited changes in time course of the PSS or TBW based on immediately prior (rapid calibration) or cumulative (cumulative recalibration) sensory history. We first sought to assess the cumulative time course of temporal recalibration in the presence and absence of feedback as participants completed the SJ task. In the presence of feedback, the PSS, as illustrated in Figure 4-2 (top left panel), decreased in absolute value over time. This shift in the PSS toward objective synchrony (i.e., an SOA of 0 ms) became significant for the interval between trials 413 and 623 (black bar denotes period of significant difference). In contrast, the absence of feedback (bottom left panel) failed to result in any significant changes in the PSS over the course of the experiment. These results illustrate that feedback coupled to the presentation of sensory information, but not sensory statistics alone, were responsible for the shift in the PSS toward objective synchrony.

With regard the TBW, as illustrated in Figure 4-2 (right panel), there was a significant narrowing for groups receiving feedback (top panel, significant between trials 170 and 720) as well as for those who did not (bottom panel, significant between trials 370 and 540). These results highlight a cumulative recalibration of audiovisual temporal acuity, even under circumstances of more passive sensory stimulation. However, the dynamics of these changes differed between the feedback and no feedback conditions, with the narrowing arising more rapidly in the presence of feedback, and persisting until the end of the feedback epoch (as opposed to the transient effect observed in the case of no feedback).

In addition to examining the effects during the training interval (i.e., the period in which feedback was given, trials 1-720), we also sought to examine the durability of these effects following the removal of feedback. This assessment was carried out over trials 721-860. For individuals previously given feedback, the earlier effect of cumulative recalibration of the TBW

persisted over trials 721-744 but dissipated with time. For the group that did not receive feedback, no change in the time course of cumulative recalibration was observed during this period (trials 721-860). For the PSS, not surprisingly we did not observe any additional changes in the time course of cumulative recalibration during this period for both the group receiving feedback and the group that did not receive feedback.

In contrast to these cumulative recalibration effects on both the PSS and TBW, little was seen in regards to a change in the time course of rapid recalibration (see Figure 4-5). Thus, when the data was analyzed on the basis of the immediately preceding trial (audio- vs. video-leading), there were no apparent changes in the time course of recalibration effects for either the PSS or the TBW, or for the feedback and no feedback groups. We did, however, observe that for the mean of all trials, the PSS (Feedback  $M = 14.8$ ,  $p < .001$ ; No Feedback  $M = 11.1$ ,  $p < .001$ ) and TBW (Feedback  $M = 5.1$ ,  $p < .001$ ; No Feedback  $M = 5.1$ ,  $p < .001$ ) were significantly shifted on a trial-to-trial basis. Hence, we conclude that while immediately prior sensory experience (i.e., bottom-up factors), shifted the PSS on a trial-to-trial basis, sensory experience alone is not sufficient to influence a change in the time course of rapid recalibration.

Collectively, the results illustrate that changes in sensory statistics alone is enough to drive perceptual learning, as defined by the cumulative narrowing of the TBW in the no feedback group. However, the time course of this plasticity is accelerated when feedback was provided. Further, changes in the PSS, which took place over a slower time scale when compared to the TBW, indicated that this measure is more stable as compared to the TBW (a finding reinforced by the lack of change for the PSS in the no feedback conditions). Although prior work has shown perceptual learning in the absence of a reinforcement signal (Fujisaki et al., 2004; Vroomen et al., 2004), the enhanced temporal recalibration observed when a feedback signal is present resembles the enhancing effect of feedback for other forms of perceptual learning (Keetels & Vroomen, 2008; Navarra et al., 2007). Indeed, as we observe for

the PSS, feedback may be essential for perceptual learning to occur under some circumstances (Watanabe, Náñez, & Sasaki, 2001; Watanabe & Sasaki, 2015).

*Positive and Negative Feedback Differentially Impact the Time Course of Temporal Recalibration*

As illustrated above, feedback strongly influences the time course of cumulative audiovisual temporal recalibration. However, how this feedback is driving these changes remains an open question. Stated differently, individuals received two forms of feedback in the context of this task – positive feedback when they were correct in their judgment and negative feedback when they were incorrect in their judgment. Do these two types of feedback differentially impact the time course of temporal recalibration? That such a distinction might exist is grounded in evidence from studies of reward system circuitry, which show that this system is differentially activated by positive and negative feedback and is underpinned by distinct neural networks (Aaron R Seitz & Dinse, 2007). Additionally, although no change was seen in rapid recalibration in the presence of feedback, this initial analysis lumped together positive and negative feedback, which may have masked differential effects based on prior feedback history. Hence, we analyzed both cumulative and rapid recalibration effects of the PSS and TBW as a function of whether individuals were correct (i.e. received positive feedback) or incorrect (i.e. received negative feedback) on the previous trial.

Findings revealed a relatively small effect of feedback type on the dynamics of PSS cumulative recalibration. Negative feedback drove a very transient change in the PSS toward true synchrony (i.e. smaller absolute value; significant between trials 392-409; Figure 4-3 top left black lines). Positive feedback did not elicit significant cumulative recalibration of the PSS (Figure 4-3, bottom left black lines). In contrast, for the TBW, cumulative recalibration was greatly impacted by feedback. Following prior incorrect responses (i.e., negative feedback), narrowing of the cumulative TBW was evident earlier and sustained over a longer time course

(trials 200-720) than changes to the TBW observed following prior correct responses (i.e., positive feedback) (trials 418-551; Figure 4-3, right). Collectively, these results support the conclusion that feedback that informs (i.e., incorrect feedback), rather than confirms (i.e., correct feedback), a perceptual decision accelerates and sustains perceptual learning.

For rapid recalibration, immediately preceding positive feedback elicited a significant change in PSS that began relatively early and lasted for the duration of the feedback (i.e., significant change in PSS between trials 239 and 720 for t-1 correct trials; Figure 4-3 bottom left red lines). Thus, it appears that following a signal confirming a perceptual decision, individuals exhibited a greater propensity for adjusting their PSS on a trial-by-trial basis. In contrast, no significant change in the PSS was seen after negative feedback. No change in rapid recalibration of the TBW was observed as a function of positive or negative feedback.

In order to examine the interaction between cumulative and rapid recalibration effects as a function of feedback type, separate 2 (cumulative vs. rapid) X 2 (previous trial correct vs. incorrect) within-subjects ANOVAs for the PSS and TBW were conducted. As illustrated in Figure 4-3, a significant interaction was observed for the PSS (trials 301-399; as illustrated by the gray shaded area). This effect was driven by the finding that when on the previous trial participants had been informed of an incorrect response, the time course of cumulative and rapid recalibration followed one another. This was not the case when the participant had been informed of a correct response on the previous trial. Hence, when participants were informed of a correct response on the preceding trial, they appear to more readily incorporate recent sensory evidence into their judgments. Summarizing these results, under conditions of informative (i.e., negative) feedback, the time courses of rapid and cumulative recalibration appear to be yoked, while under conditions of confirmative (i.e., positive) feedback, rapid and cumulative recalibration effects appear to uncouple. This uncoupling may be adaptive in that only corrective signals are able to drive rapid plasticity.

### *Time Course of Rapid Recalibration of the TBW Diverges as a Function of Prior Feedback Reliability*

To better understand the contribution of feedback to recalibration processes, and the interrelationship between rapid and cumulative recalibration effects, we tested whether changing the reliability of the feedback would differentially alter rapid vs. cumulative temporal recalibration. Prior studies of visual perceptual learning have demonstrated that while feedback enhances perceptual learning, presenting feedback that is uncorrelated to responses (i.e. unreliable feedback) impairs perceptual learning (Michael H Herzog & Fahle, 1997). Feedback was provided to different groups of participants and that was 100%, 80%, or 50% reliable (see Methods). We hypothesized that if the time course of temporal recalibration was dependent on external reinforcement, we would see progressively less temporal recalibration as feedback reliability decreased. Indeed, unlike the group receiving 100% reliable feedback (described above), we did not observe cumulative recalibration of the PSS for the groups receiving 80% or 50% reliable feedback (see Figure 4-6). In contrast however, we did observe cumulative recalibration of the TBW for all groups, although these changes were seen over a shorter extent of trials when compared with the 100% reliable feedback group (see Figure 4-6).

In order to examine the different time courses of multisensory temporal recalibration as a function of feedback reliability, we conducted separate 2 (type of recalibration: cumulative vs. rapid) X 3 (feedback reliability: 100%, 80%, 50%) between-subjects ANOVAs for the PSS and TBW along the time-series of trials such that an effect was interpreted as significant at  $\alpha < 0.01$  for at least 10 consecutive trials. For the PSS, we did not observe a main effect of type of recalibration, feedback reliability, or an interaction (see Figure 4-6). In contrast, for the TBW, we observed a significant main effect of type of recalibration between trials 103 and 841 (all  $p < 0.01$ ) and a significant type of recalibration X feedback reliability interaction between trials 816 and 844 (Figure 4-4; all  $p < 0.01$ ; indicated by the gray shading). Thus, and as is evident in

Figure 4-4, although the dynamics of cumulative temporal recalibration of the TBW failed to differ dependent upon feedback reliability (top panel; one-way between-subjects ANOVA, all  $p > 0.06$ ), the dynamics of rapid temporal recalibration of the TBW did diverge (bottom panel). Specifically, a one-way between-subject ANOVA on the rapid recalibration values demonstrated a significant effect between trials 806 and 851 (all  $p < 0.01$ ). Subsequent post-hoc t-tests performed on the rapid recalibration patterns as a function of feedback reliability demonstrated that the 50% reliable feedback elicited a higher degree of rapid recalibration (variability on a trial-by-trial basis, weighting more heavily immediately preceding sensory experience) than the 80% reliable feedback (between trials 780 and 861) and the 100% reliable feedback (between trials 801 and 827). The 100% and 80% reliable feedback conditions did not differ from one other (all  $p > .33$ ). The increase in rapid recalibration after the 50% reliable feedback signal is removed may represent an increased tendency for the subjects to disregard feedback and more heavily weigh sensory statistics when prior feedback has been unreliable in signaling the correctness of their judgments.

This finding represents the second example in our data of an uncoupling between cumulative and rapid recalibration (the first being that brought about by the correct vs. incorrect nature of the feedback). Namely, we observe that when feedback reliability is reduced, perceptual learning occurs, but with differing dynamics for cumulative and rapid recalibration, again suggesting differing mechanistic processes. We hypothesize that as a result of the conflict between sensory evidence and feedback signals, those individuals presented with the least reliable feedback (50%) were more likely to rely on immediate sensory information to recalibrate their audiovisual temporal representation. This may be due to a decreased reliance on top-down signals generated by sensory feedback and an increased reliance on bottom-up sensory information. In the groups receiving unreliable feedback, as some of the feedback was misinformative, increased reliance on sensory statistics would be adaptive in that sensory driven recalibration would produce a more accurate perceptual representation.

## General Discussion

Here we show that top-down factors (i.e., feedback signals) can interact with bottom-up signals in order to change the dynamic time course of temporal recalibration for two measures of audiovisual temporal perception (PSS and TBW). By employing a sliding window analysis for this study, we were able to characterize, for the first time, how rapid and cumulative temporal recalibration occur in both the presence and absence of feedback, and to characterize the differing temporal dynamics for these two time scales of perceptual learning. Our findings illustrate that while sensory experience alone is sufficient to elicit some degree of temporal recalibration, feedback signals can work in conjunction with sensory experience to produce greater perceptual plasticity.

That feedback signals alter the dynamics of temporal recalibration is not surprising as enhanced plasticity would be adaptive in response to changing environmental statistics or task demands. Despite this assumption, it is interesting that feedback is sufficient, if only transiently, to alter perceptual representations for which a strong history of sensory experience exists. The PSS, a measure that is reflective of an individual's internal representation of the temporal statistical structure of the external world, is rarely at true synchrony (i.e., 0 ms). Rather, this measure is typically biased toward an asynchrony in which the visual stimulus leads the auditory stimulus - reflective of the typical statistical structure of audiovisual stimuli within our world (Law & Gold, 2009; Polley, Steinberg, & Merzenich, 2006; Powers et al., 2009; Aaron R Seitz et al., 2009). Although adaptation studies have shown that repeated presentation of asynchronous audiovisual stimuli (i.e. toward either a visual or auditory leading stimulus set) can shift the PSS in the direction of the experienced asynchrony (Bischoff-Grethe, Hazeltine, Bergren, Ivry, & Grafton, 2009; Lempert & Tricomi, 2015; Liu, Hairston, Schrier, & Fan, 2011), we report a shift in the PSS in the absence of any changes in the temporal structure of the stimuli and based solely on the presence of feedback. Indeed, the changes elicited under such circumstances are

invariably in the direction of true synchrony. As we did not introduce a change in the temporal structure of the stimuli that would favor a directional shift in the PSS, we conclude that this change is driven largely by top-down factors linked to the delivery of feedback.

That the changes in TBW and PSS in response to feedback are quick to develop is also not surprising as this too may be adaptive. Interestingly, it also appears that over the course of a single session, both sensory and feedback-induced changes in the PSS and TBW can be quick to dissipate as, with the exception of the group receiving 100% reliable feedback, the time course returns to the level of the initial estimate within a relatively small number of trials after feedback is removed. As studies of perceptual training have reported changes in temporal acuity between training sessions (Dixon & Spitz, 1980; Zampini, Shore, & Spence, 2003), it is possible that by extending our analysis across multiple sessions we might observe further changes in the time course of recalibration. Future investigations will be necessary to determine if sensory experience or feedback elicit durable changes in the PSS and TBW or whether the plasticity we observe is simply reflective of fast adaptation.

Future studies to explore the time course of temporal recalibration would also determine if unreliable feedback elicits lasting changes beyond the post-feedback period measured in this study that differ from any changes elicited by sensory experience in the absence of feedback. Although we observed similarities in the time course of recalibration of the TBW and PSS in the absence of feedback and with unreliable feedback, the mechanism by which these changes occur may be different. It is possible that the return to original levels in performance we observed may be due to a change in criteria when unreliable feedback was present (i.e. trials 1-720) rather than perceptual learning that results in a lasting change in the perceptual representation (M. H. Herzog & Fahle, 1999). Changing perceptual decision criteria in response to erroneous (i.e. unreliable) feedback has been suggested to be adaptive as such a transient change in criteria would minimize error signals while protecting prior representations of the stimuli (Michael H. Herzog, Ewald, Hermens, & Fahle, 2006). Thus, when the unreliable

feedback signal is removed, the criterion can be rapidly adjusted to criterion prior to exposure to unreliable feedback. Although changes in criterion are usually limited to a perceptual training session, changes in criterion are typically not observed during a second session a day later whereas changes in sensitivity are maintained after at least a day (Aberg & Herzog, 2012). By extending the time course analysis of temporal recalibration beyond a single day, we would hypothesize that if unreliable feedback elicited any durable change in the PSS or TBW, we would observe a change in the PSS or TBW relative to the final estimate of the TBW or PSS on the first day.

Increasing evidence suggests that the mechanisms supporting unisensory (i.e., within-modality) perceptual learning are evident at higher cortical levels (Watanabe & Sasaki, 2015; Zilber, Ciuciu, Gramfort, Azizi, & Van Wassenhove, 2014) and that enhanced perception of amodal sensory properties due to perceptual training in modality can exhibit transfer across sensory modalities to an untrained sensory modality (Barakat, Seitz, & Shams, 2015; McGovern, Astle, Clavin, & Newell, 2016). Stimulus exposure that is more passive in nature appears to drive changes at lower cortical levels while increasingly the relevance of the stimulus properties elicits changes at both higher and lower cortical levels (Watanabe et al., 2002). Multisensory stimuli, which engage a larger cortical network than unisensory stimuli, may facilitate perceptual learning by increasing activity of primary sensory regions as well as higher level sensory cortex. Recent evidence suggests that multisensory interactions, while present across different levels of the cortical hierarchy, may differ in their computational functions across higher-order and sensory regions (Fujisaki et al., 2004; Vroomen et al., 2004). Accordingly, a feedback signal may also serve to engage a larger cortical network, which in turn enables a greater capacity for perceptual learning to occur.

A possible explanation for why greater temporal recalibration occurs with feedback, provided it is reliable, is that sensory readout is improved for higher order cortical areas involved in sensory decision-making due to the feedback signal. At the neural level, this is in line with

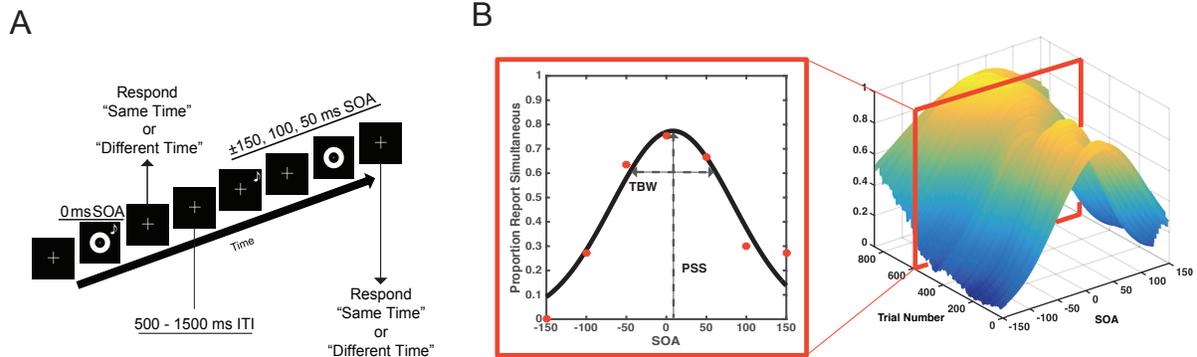
studies of visual perceptual learning that observed changes in activity patterns in the anterior cingulate cortex to track changes in decision making during visual perceptual learning (Powers et al., 2012; Powers et al., 2009). Furthermore, neural evidence suggests that prediction error signals during perceptual learning refine and strengthen neural connectivity between sensory neurons and those neurons required for the perceptual response and thus may support changes in higher-order regions (Rohe & Noppeney, 2016). Thus, in the absence of an informative reinforcement signal, rapid but transient changes in perceptual plasticity are likely due to changes in low-level sensory areas. Future investigations will be necessary to determine if changes in the connectivity of higher-order cortical areas and low-level sensory processes underlie the observed changes in temporal recalibration and if these changes are durable or transient.

## **Conclusions**

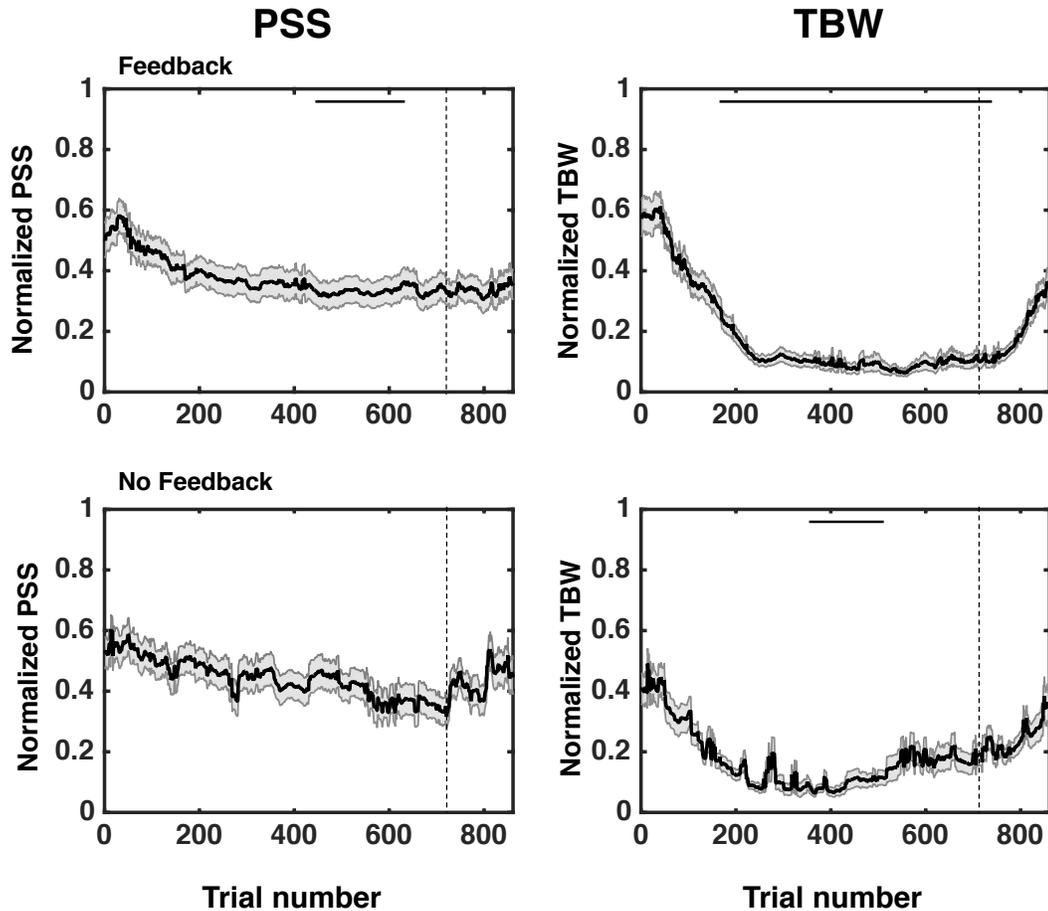
We report that sensory experience and feedback signal interact to drive both rapid and cumulative temporal recalibration of the TBW and PSS for audiovisual stimuli. While rapid and cumulative temporal recalibration often follow similar time courses, these time courses may diverge dependent upon prior feedback signals. Our findings support that prior sensory history feedback signals influence subsequent perceptual plasticity to elicit both rapid and cumulative temporal recalibration.

## **Acknowledgements**

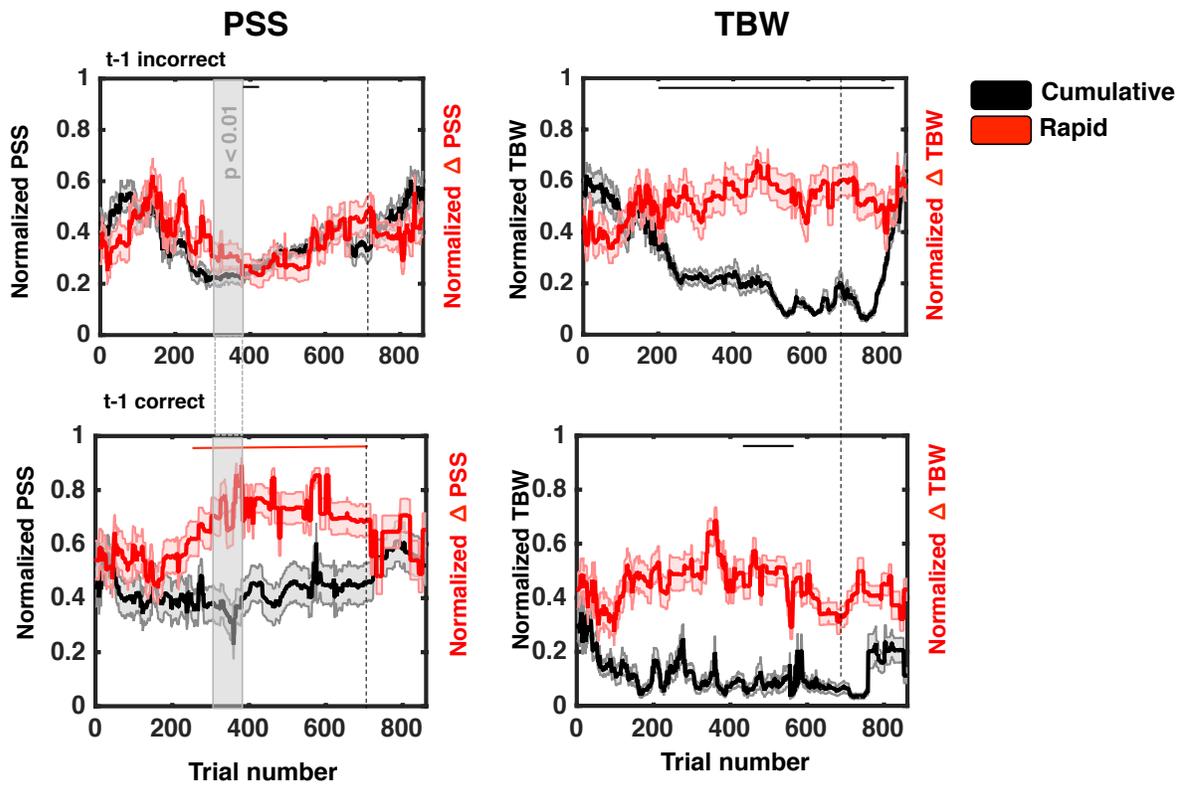
The project was supported by NIH grants CA183492 and HD083211, the Simons Foundation Autism Research Initiative and the Wallace Foundation. JPN was additionally supported by a NSF GRF (NSF-14-590). The authors declare no conflicts of interest. Data from this study can be viewed at <https://vanderbilt.box.com/s/a04pr5dbsjrhrnquz5lrcvv5i6buafdp>.



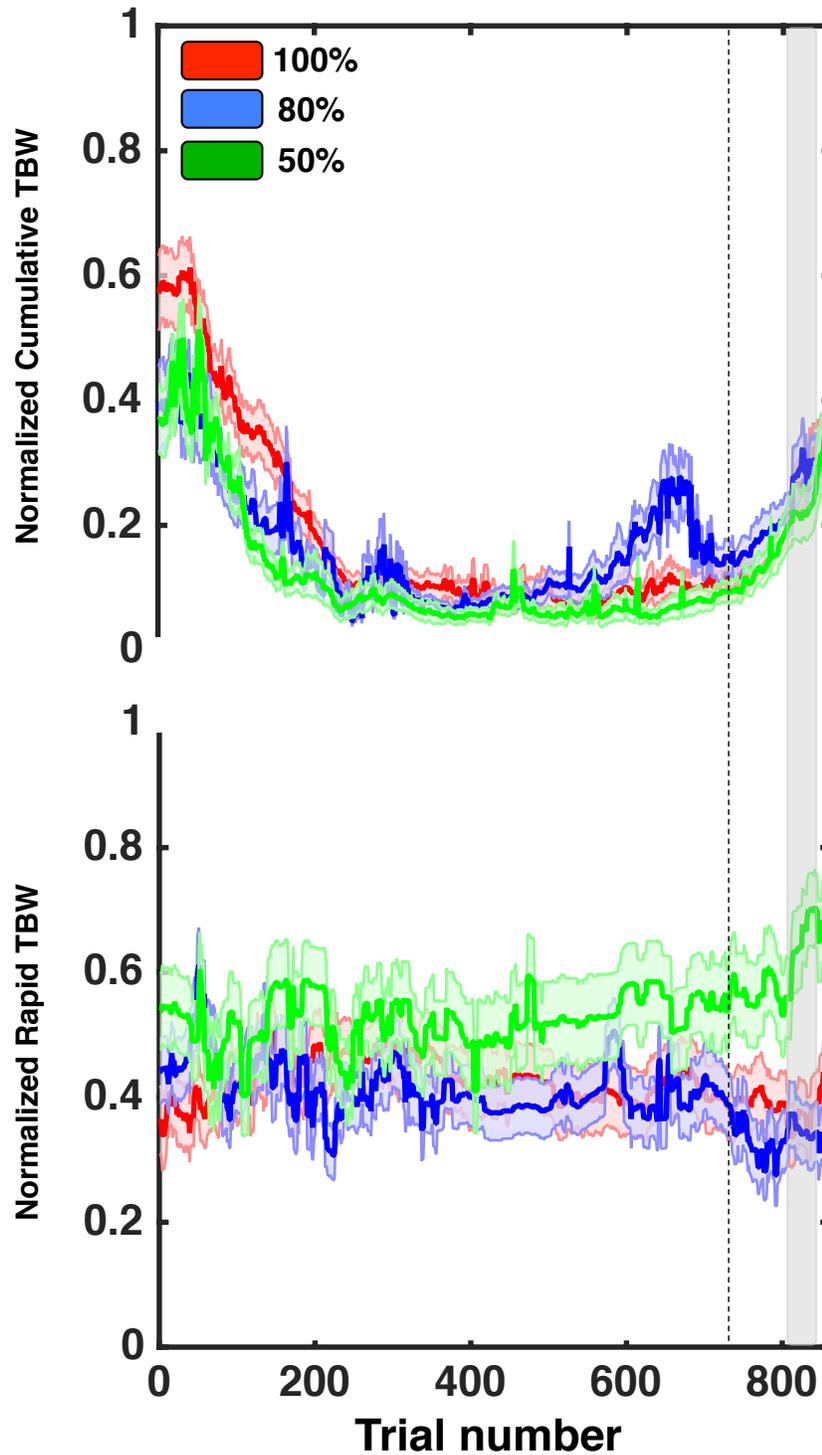
**Figure 4-1. Measurement of Cumulative and Rapid Temporal Recalibration** A) Representation of a trial sequence for the simultaneity judgment (SJ) task. Participants were asked to judge if stimuli occurred at the same time or different times. B) Individual fittings for a single participant using the sliding window approach across Trials 1-860. The inset on the left shows a single fitting at one time-point along the time course with the PSS (mean of distribution) and TBW (standard deviation of distribution) at that particular moment in time. The TBW, PSS (cumulative), and  $\Delta$ TBW,  $\Delta$ PSS (rapid), were normalized on a within-subject basis, and in order to correct for multiple comparisons we consider an effect significant at  $\alpha < 0.01$  for at least 10 consecutive trials. Trial 0 was defined as the 140 trials utilized to establish initial estimates of the PSS and TBW. The time course analysis was conducted on the following 860 trials. From Trial 1-720, participants were assigned to one of four groups that received varying amounts of feedback following a response. From Trials 721-860, no feedback was presented following a response for all participants.



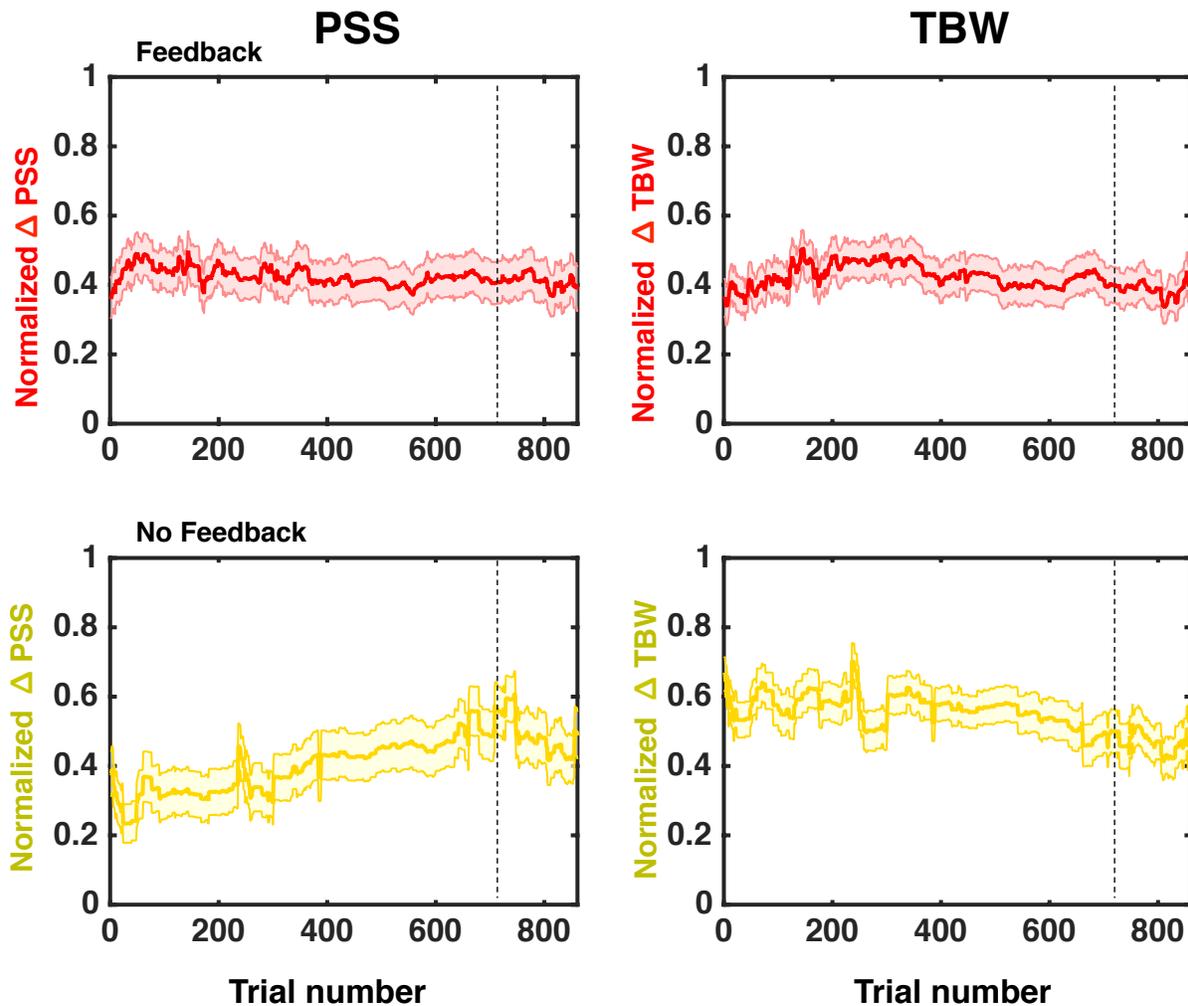
**Figure 4-2. Cumulative Temporal Recalibration** The time course of cumulative (grayscale) multisensory temporal recalibration of the PSS (left) and TBW (right) with (upper row) and without feedback (lower row). Solid bars shown above the time course are indicative of at least 10 consecutive trials at which the PSS or TBW (cumulative recalibration) significantly differed from Trial 0 ( $\alpha < 0.01$  for all trials). Shaded region illustrates S.E.M. at each trial across the time course analysis. The post-feedback period beginning at trial 721 is denoted by the gray, dashed line.



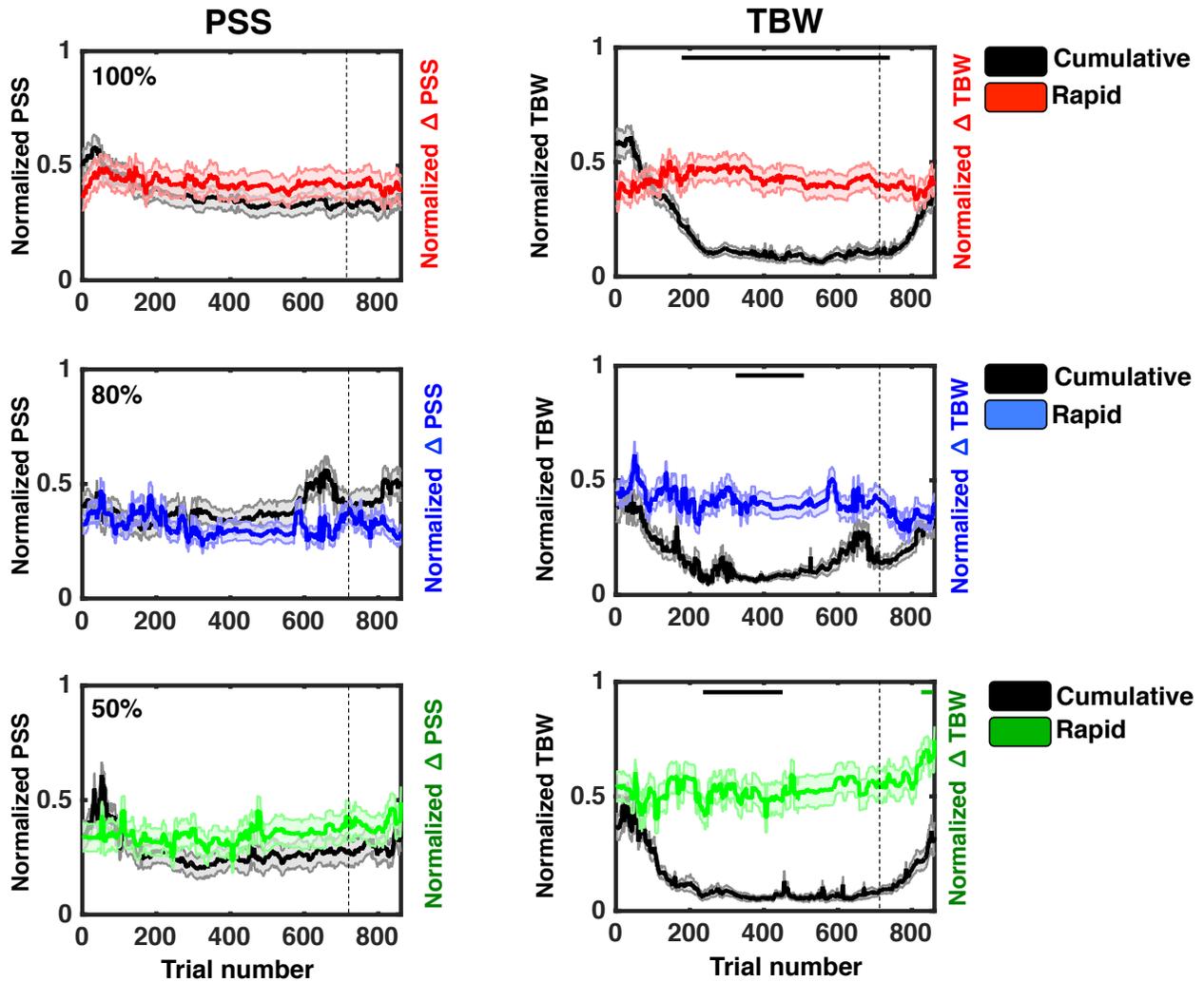
**Figure 4-3. Effect of Prior Feedback on Temporal Recalibration** The time course of multisensory rapid (color) and cumulative (grayscale) temporal recalibration as a function of prior negative (upper row) and positive (lower row) feedback on trial t-1. Solid bars shown above the time course are indicative of at least 10 consecutive trials at which the PSS or TBW (cumulative) or  $\Delta$ PSS or  $\Delta$ TBW (rapid) significantly differed from Trial 0 ( $\alpha < 0.01$  for all trials). The trials for which a significant interaction of temporal recalibration (cumulative vs. recalibration) X feedback (t-1 correct vs. t-1 incorrect) indicated by solid gray shading ( $\alpha < 0.01$  for all trials).



**Figure 4-4. Effect of False Feedback on the Time Course of Temporal Recalibration** The time course of cumulative (upper) and rapid (lower) temporal recalibration of the TBW for feedback that was 100% (red), 80% (blue), and 50% (green) reliable. We observe that the time course of rapid recalibration of the TBW for the group receiving uncorrelated feedback signals (50% reliable) diverges from the other group in the post-feedback trial block such that there is greater trial-to-trial readjustment of the TBW (gray shading indicates interaction).



**Figure 4-5. Rapid Temporal Recalibration** The time course of rapid temporal recalibration in the presence and absence of feedback. The normalized magnitude of trial-to-trial change (i.e. rapid recalibration) in the PSS (left) and TBW (right) did not change over the course of the second trial block (Trials 1-720) when feedback was either present (red) or absent (yellow). Similarly, the magnitude of rapid recalibration remained stable when feedback was removed for the one group during the third trial block (Trials 721-860; right of dashed line). While we did not observe a change in the magnitude of rapid recalibration, the PSS (Feedback  $M = 14.8$ ,  $p < .001$ ; No Feedback  $M = 11.1$ ,  $p < .001$ ) and TBW (Feedback  $M = 5.1$ ,  $p < .001$ ; No Feedback  $M = 5.1$ ,  $p < .001$ ) were significantly shifted trial-to-trial consistent with prior reports (Van der Burg et al., 2013; Van der Burg, Orchard-Mills, et al., 2015) of the overall rapid temporal recalibration observed over the course many trials. This analysis corresponds to the time course of cumulative temporal recalibration in Figure 2.



**Figure 4-6. The Effect of False Feedback on Cumulative and Rapid Temporal Recalibration** The time course of cumulative (gray scale) and rapid recalibration (color) for participants receiving 100% (red), 80% (blue), and 50% (green) feedback during the second trial block (Trials 1-720) and when feedback is removed (Trials 721-860). We did not observe cumulative recalibration of the PSS (left) for either group receiving unreliable (80% or 50% reliable; middle and bottom row) feedback. This lack of change in cumulative recalibration for the PSS is comparable to the effects seen in individuals without access to feedback, but differs strikingly from those in which the feedback was 100% reliable (top row). With regard the TBW (right), both groups receiving 80% and 50% reliable feedback exhibited transient cumulative temporal recalibration resulting in significant narrowing of the TBW (Trials 305-541 and Trials 219-430 respectively) which was unlike the time course exhibited by those for which feedback was reliable on 100% of trials (Trials 170 to 720). Thus, the time course of temporal recalibration when feedback is less reliable resembles the time course observed when feedback is absent. A significant change in the magnitude of rapid recalibration was observed only for the group receiving 50% reliable feedback during the second trial block (green) after the unreliable feedback signal was removed (Trials 808-860). No change in the magnitude of rapid recalibration was observed for the groups receiving either 100% or 80% reliable feedback. Solid bars shown above the time course are indicative of at least 10 consecutive trials at which the PSS or TBW (cumulative) or  $\Delta$ PSS or  $\Delta$ TBW (rapid) significantly differed from Trial 0 ( $\alpha < 0.01$  for all trials).

## References

- Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision, 12*(3). doi:10.1167/12.3.3
- Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology, 25*(2), R60-R61.  
doi:10.1016/j.cub.2014.12.011
- Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R. B., & Grafton, S. T. (2009). The influence of feedback valence in associative learning. *NeuroImage, 44*(1), 243-251.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.
- Bruns, P., & Röder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past. *Scientific reports, 5*.
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.
- Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Current opinion in neurobiology, 15*(2), 154-160.
- Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision research, 33*(3), 397-412.
- Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision research, 35*(21), 3003-3013.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature neuroscience, 7*(7), 773-778.
- Ghose, G. M. (2004). Learning in mammalian sensory cortex. *Current opinion in neurobiology, 14*(4), 513-518.
- Harvey, C., Van der Burg, E., & Alais, D. (2014). Rapid temporal recalibration occurs crossmodally without stimulus specificity but is absent unimodally. *Brain Research, 1585*, 120-130.

- Herzog, M. H., Ewald, K. R. F., Hermens, F., & Fahle, M. (2006). Reverse feedback induces position and orientation specific changes. *Vision research*, *46*(22), 3761-3770.  
doi:10.1016/j.visres.2006.04.024
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision research*, *37*(15), 2133-2141.
- Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision research*, *39*(25), 4232-4243. doi:10.1016/s0042-6989(99)00138-8
- Keetels, M., & Vroomen, J. (2007). No effect of auditory–visual spatial disparity on temporal recalibration. *Experimental Brain Research*, *182*(4), 559-565.
- Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile–visual asynchronous stimuli. *Neuroscience letters*, *430*(2), 130-134.
- Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat Neurosci*, *11*(4), 505-513.  
doi:[http://www.nature.com/neuro/journal/v11/n4/suppinfo/nn2070\\_S1.html](http://www.nature.com/neuro/journal/v11/n4/suppinfo/nn2070_S1.html)
- Law, C.-T., & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature neuroscience*, *12*(5), 655-663.
- Lempert, K. M., & Tricomi, E. (2015). The Value of Being Wrong: Intermittent Feedback Delivery Alters the Striatal Response to Negative Feedback. *Journal of Cognitive Neuroscience*.
- Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1219-1236.
- McGovern, D. P., Astle, A. T., Clavin, S. L., & Newell, F. N. (2016). Task-specific transfer of perceptual learning across sensory modalities. *Current Biology*, *26*(1), R20-R21.
- Medina, J., & Rapp, B. (2014). Rapid experience-dependent plasticity following somatosensory damage. *Current Biology*, *24*(6), 677-680.

- Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience letters*, 413(1), 72-76.
- Noel, J.-P., Łukowska, M., Wallace, M., & Serino, A. (2016). Multisensory simultaneity judgment and proximity to the body. *Journal of Vision*, 16(3), 21-21.
- Noel, J.-P., Wallace, M. T., Orchard-Mills, E., Alais, D., & Van der Burg, E. (2015). True and Perceived Synchrony are Preferentially Associated With Particular Sensory Pairings. *Scientific reports*, 5.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4), 437-442.
- Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *The Journal of Neuroscience*, 26(18), 4970-4982.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, 32(18), 6263-6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, 29(39), 12265-12274.
- Powers III, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports*, 6, 23374.
- Rohe, T., & Noppeney, U. (2016). Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. *Current Biology*.
- Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current opinion in neurobiology*, 17(2), 148-153.
- Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron*, 61(5), 700-707.

- Seitz, A. R., Nanez, J. E., Sr., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, 6(9), 966-973. doi:10.1167/6.9.9
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, 12(11), 411-417.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15(8), 520-535.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, 227(2), 249-261. doi:10.1007/s00221-013-3507-3
- Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid Recalibration to Audiovisual Asynchrony. *Journal of Neuroscience*, 33(37), 14633-14637. doi:10.1523/jneurosci.1182-13.2013
- Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific reports*, 5.
- Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B-Biological Sciences*, 282(1804). doi:10.1098/rspb.2014.3083
- Van der Burg, E., Orchard-Mills, E., & Alais, D. (2015). Rapid temporal recalibration is unique to audiovisual stimuli. *Experimental Brain Research*, 233(1), 53-59.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, 72(4), 871-884.
- Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive brain research*, 22(1), 32-35.

- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105-123.
- Watanabe, T., Náñez, J. E., 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. *Nature neuroscience*, *5*(10), 1003-1009.
- Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*(6858), 844-848.
- Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: toward a comprehensive theory. *Annual review of psychology*, *66*, 197.
- Wozny, D. R., & Shams, L. (2011). Recalibration of Auditory Space following Milliseconds of Cross-Modal Discrepancy. *Journal of Neuroscience*, *31*(12), 4607-4612.  
doi:10.1523/jneurosci.6079-10.2011
- Yu, L., Stein, B. E., & Rowland, B. A. (2009). Adult plasticity in multisensory neurons: short-term experience-dependent changes in the superior colliculus. *The Journal of Neuroscience*, *29*(50), 15910-15922.
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*(2), 198-210.
- Zilber, N., Ciuciu, P., Gramfort, A., Azizi, L., & Van Wassenhove, V. (2014). Supramodal processing optimizes visual perceptual learning and plasticity. *NeuroImage*, *93*, 32-46.

## Chapter V

### Perceptual Training Enhances Temporal Acuity for Multisensory Speech

#### Abstract

The temporal relationship between auditory and visual cues is essential for determining if these signals will be integrated. The temporal binding window (TBW) is a construct that describes the epoch of time during which asynchronous auditory and visual stimuli are perceptually bound. Recently, a number of studies have demonstrated the capacity for perceptual training to enhance temporal acuity for audiovisual stimuli (i.e., narrow the TBW). These studies, however, have only examined multisensory perceptual learning that develops in response to feedback that is provided when making judgments on simple, low-level audiovisual stimuli (i.e. flashes and beeps). Here we sought to determine if perceptual training was capable of altering temporal acuity for audiovisual speech. Furthermore, we also explored whether perceptual training with simple or complex audiovisual stimuli generalized across levels of stimulus complexity. Using a simultaneity judgment (SJ) task, we measured individuals' temporal acuity (as estimated by the TBW) prior to, immediately following, and one week after four consecutive days of perceptual training. We report that while enhanced temporal acuity for audiovisual speech stimuli is observed immediately following perceptual training, the observed changes in temporal acuity are not as durable as those seen for simple stimuli. Additionally, we find that changes in temporal acuity following perceptual training do not generalize across the levels of stimulus complexity in this study. Overall, the results suggest that perceptual training is capable of enhancing temporal acuity for audiovisual speech in adults.

## Introduction

The capacity to integrate sensory information from different sensory modalities is essential for constructing a coherent perception of the world (Stein & Stanford, 2008; Wallace & Stevenson, 2014). Combining sensory cues from different modalities, termed multisensory integration, often results in improved behavioral performance such as speeded reaction times (Diederich & Colonius, 2004), enhanced localization (Zweig, Suzuki, & Grabowecky, 2015) and detection (Lovelace, Stein, & Wallace, 2003), and can result in a host of illusory percepts (McGurk & MacDonald, 1976; Shams, Kamitani, & Shimojo, 2000). The behavioral benefits of multisensory integration are notably evident in our everyday lives when conversing as the integration of auditory and visual cues enhance our comprehension of speech stimuli (Grant & Seitz, 2000; Schwartz, Berthommier, & Savariaux, 2004), particularly in noisy environments (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumbly & Pollack, 1954). Thus, when an individual is able to view the dynamic facial movements produced by the articulation of speech with the associated auditory signal, comprehension is enhanced (Jaekl, Pesquita, Alsius, Munhall, & Soto-Faraco, 2015; MacLeod & Summerfield, 1987). The McGurk effect has been used as a powerful tool to index the perceptual binding of auditory and visual speech signals. In the most common incarnation of this perceptual illusion, the artificial pairing of the auditory token /ba/ with the visual token /ga/ frequently results in the perception of a novel percept, generally /da/ or /tha/ (McGurk & MacDonald, 1976) further illustrates how the combination of these visual and auditory sensory cues allows the brain makes top-down inferences as the expected cues in each corresponding modality (Gau & Noppeney, 2016; Olasagasti, Bouton, & Giraud, 2015). At the neural level, multisensory interactions in response to speech are readily evident as visual speech cues enhance the processing of auditory speech signals (van Wassenhove, Grant, & Poeppel, 2005), and audiovisual speech signals result in enhanced activity in both multisensory brain regions [18] as well as those involved in speech production (Venezia et al., 2016).

The integration of multisensory stimuli, including audiovisual speech cues, is highly dependent upon the temporal relationship of the respective unisensory cues, with more temporally proximate stimuli have a greater likelihood of being integrated (Meredith, Nemitz, & Stein, 1987; van Wassenhove, Grant, & Poeppel, 2007). From these observations, the concept of the temporal binding window (TBW) has emerged to refer to the epoch of time during which physically asynchronous stimuli are likely to be integrated and perceptually bound (Wallace & Stevenson, 2014). Conceptually, the TBW has a great deal of ecological validity as the brain must constantly resolve asynchrony of auditory and visual signals which differ in physical transmission times (i.e. due to varying distances) (Vroomen & Keetels, 2010). The TBW is not a uniform construct but rather is highly dependent upon the nature of the stimuli that are to be combined. Thus, the smallest TBWs are typically observed for the pairing of simple, static stimuli (i.e. flashes and beeps), and increasingly larger TBWs are seen for pairings of more complex stimuli (i.e. biological motion and speech) (Dixon & Spitz, 1980; R. A. Stevenson & Wallace, 2013; Vatakis & Spence, 2006a). Nonetheless, within an individual, there is a high degree of correlation in the size of these different windows, suggesting a common neural architecture subserving this form of perceptual integration (R. A. Stevenson & Wallace, 2013). Additionally, over the course of typical development, the TBW narrows, but does so over a surprisingly long maturational timeline, not fully maturing until late adolescence. Finally, as temporal acuity for multisensory stimuli has been reported to be atypical for a number of developmental disorders (see (Wallace & Stevenson, 2014) for review), including autism spectrum disorder, there has been an increasing interest in better understanding multisensory temporal processing, particularly for speech stimuli.

As for all audiovisual stimuli, the temporal relationship of auditory and visual speech cues is an essential factor that influences the ultimate percept (Summerfield, 1992). As the facial movements of speech production are highly correlated with the temporal features of the auditory speech signal (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009), it

is likely that many of the behavioral benefits derived from the combined use of auditory and visual information is a result of their close temporal proximity. This relationship was nicely demonstrated by *van Wassenhove et al. 2007* who showed that the magnitude of multisensory integration (as measured by the strength of the McGurk) is greatest when the auditory and visual cues are approximately synchronous or slightly visual-leading, which parallels the peak of the TBW when measured for the same stimuli (*van Wassenhove et al., 2007*). Indeed, the strength of multisensory integration has also been observed to correlate with individual audiovisual temporal acuity, as decreased temporal acuity is reported to be directly proportional to decreased perception of the McGurk effect (*Ryan A. Stevenson, Zemtsov, & Wallace, 2012*). Furthermore, and of more ecological relevance, the capacity to use auditory and visual cues to better comprehend audiovisual speech is reported to be enhanced for individuals with greater audiovisual temporal acuity (*Conrey & Pisoni, 2006*). Collectively, such findings suggest that the acuity of a general audiovisual temporal mechanism may predict the amount of behavioral benefit one derives from these cues as well as more generally language and communication abilities. As individuals with autism spectrum disorder (ASD) have recently been shown to have wider TBWs for audiovisual speech stimuli (*Ryan A Stevenson, Siemann, Schneider, et al., 2014*) compared to typically developing peers, interventions that target improvements in audiovisual temporal acuity might also lead to more generalized improvements in speech perception.

Although numerous early studies of perceptual learning reported that perceptual training enhances performance within a signal sensory modality (*Herzog & Fahle, 1997; Aaron R Seitz & Dinse, 2007; Aaron R. Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006*), there has been an increasing interest in exploring perceptual plasticity for multisensory stimuli (*Shams & Seitz, 2008*). Recently, it has been reported by *Powers et al. 2009* that by employing a perceptual training paradigm, temporal acuity for simple “flash-beep” stimuli could be enhanced in adults (*Powers, Hillock, & Wallace, 2009*). These findings not only demonstrated that the

TBW remains plastic in adulthood, but also showed that the training-related effects on temporal acuity were durable, still being evident a week after the cessation of training. Subsequent studies using similar stimuli have demonstrated that analogous perceptual training paradigms are capable of eliciting generalized improvements in multisensory (Setti et al., 2014) and visual temporal acuity (Powers III, Hillock-Dunn, & Wallace, 2016) as well as changes in connectivity between multisensory and primary sensory cortical areas (Powers, Hevey, & Wallace, 2012). Despite this growing interest in multisensory perceptual learning, no study has yet to employ these perceptual training paradigms to enhance temporal acuity for the most naturalistic of multisensory stimuli – audiovisual speech. In addition, no study to date has definitively established whether training at one level of stimulus complexity generalizes to a different level of stimulus complexity. The presence of such generalization has implications for both neural architecture and for the practical application of perceptual plasticity-based remediation approaches.

Here, we sought determine if temporal acuity (as indexed by changes in the TBW) for audiovisual speech could be enhanced following perceptual training using a similar paradigm as that employed by *Powers et al. 2009* and others (Powers et al., 2012; Ryan A. Stevenson, Wilson, Powers, & Wallace, 2013). Furthermore, we examined if perceptual training that narrowed the TBW using either simple, low-level “flash-beep” stimuli or more complex audiovisual speech stimuli generalized across different levels of stimulus complexity. We hypothesized that if perceptual training enhanced multisensory temporal acuity through a common temporal processing mechanism, then the effects of training at one stimulus complexity would transfer across levels (either up, down, or both). We report that while perceptual training can be structured so as to enhance temporal acuity for both simple “flash-beep” and more complex speech stimuli, these improvements do not generalize across the different levels of stimulus complexity employed by this study.

## **Methods**

### *Subjects*

Prior to participation in the study, all participants received written informed consent. A total of  $N = 28$  adult individuals (Age,  $M = 20.61$  years, Gender, 14 female) participated in the study. Participants were randomly assigned to one of three groups: low-level “flash-beep” training ( $n = 8$ ), audiovisual speech training ( $n = 11$ ), or audiovisual speech exposure ( $n = 9$ ). An additional three participants were enrolled in the study but were excluded from final analyses as they did not meet a minimal performance criteria for the SJ task (participants were excluded if they did not perceive at least  $\geq 80\%$  of the trials to be synchronous for at least one SOA across all SOAs measured). All participants had self-reported normal hearing and normal or corrected to normal vision. All recruitment and experimental procedures were approved by the Vanderbilt University Institutional Review Board and in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### *Stimuli*

All visual stimuli were presented at approximately 48 cm from the participants on a monitor (Samsung syncmaster 22 inch 2233 RZ LCD) with a refresh-rate of 60 Hz. Auditory stimuli were presented binaurally via headphones (Sennheiser HD 558) with no interaural time or level differences. The intensity of all auditory stimuli was calibrated using a sound level meter (Larson Davis SoundTrack® LxT2). For all tasks, participants were seated in a light and sound attenuating WhisperRoom™ (SE 2000 Series, Whisper Room Inc.)

For the low-level “flash-beep” SJ task, the visual stimulus consisted of a white ring on a black background that subtended  $7.2^\circ$  of visual space with an outer diameter of 6.0 cm and an inner diameter of 3.0 cm. The “flash” was presented for 16.6 ms (the duration of a single screen refresh cycle). The auditory “beep” consisted of a 10 ms, 1800 Hz tone that was presented at 83

dB. MATLAB (The MathWorks, Inc.) with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) was used to present the “flash-beep” stimuli.

For audiovisual speech SJ task, the stimulus consisted of a video of a female speaker uttering the phonemic unit /ba/ that has been employed by numerous prior studies (Quinto, Thompson, Russo, & Trehub, 2010; R. A. Stevenson & Wallace, 2013). The video spanned an area of 720 x 480 pixels and was presented in color. The duration of the video was 1833 ms and included all pre-articulatory gestures (See Figure 5-1A). The same video was utilized for perceptual training sessions for the audiovisual speech training group and the audiovisual speech exposure group, although on some trials for the exposure group a version of the video was presented with a red tint but identical to the color version in all other aspects (see explanation of audiovisual speech exposure task). All speech stimuli for SJ tasks were presented using E-Prime version 2.0.8.

### *Simultaneity Judgment Tasks*

For the “flash-beep” SJ task, participants were instructed to judge whether the visual stimulus and auditory stimulus “were synchronous, at the same time” or “were asynchronous, at different times” by pressing either 1 or 2 respectively using a keyboard. A fixation marker (1 cm x 1 cm) on a black background was present on the screen both between trials and throughout the duration of a trial including presentation of the visual stimulus. Participants were asked to maintain fixation on the fixation marker throughout the experiment. Each trial consisted of the visual and auditory stimuli presented at true synchrony (0 ms of asynchrony) or at one of the following SOAs:  $\pm 400$ ,  $\pm 300$ ,  $\pm 250$ ,  $\pm 200$ ,  $\pm 150$ ,  $\pm 100$ , and  $\pm 50$  (negative indicates the auditory stimulus was leading while positive indicates the visual stimulus was leading). SOAs for all tasks were verified externally using an oscilloscope. Following presentation of visual and auditory stimuli, a response screen was presented at which time subjects could make a response. The inter trial interval (ITI) between trials was randomly jittered from 500 to 1500 ms.

For the audiovisual speech SJ task, participants were instructed to judge whether the auditory and visual stimuli were “synchronous...properly lined up” or “asynchronous...were NOT properly lined up” by pressing 1 or 2 respectively using a keyboard. A fixation marker (1 cm x 1 cm) was present during the ITI but not during the presentation of the video. Participants were asked to maintain fixation on the fixation marker when present throughout the experiment. Each trial consisted of the video presented at true synchrony (0 ms of asynchrony) or at one of the following SOAs:  $\pm 400$ ,  $\pm 300$ ,  $\pm 250$ ,  $\pm 200$ ,  $\pm 150$ ,  $\pm 100$ , and  $\pm 50$  (the same SOAs as the “flash-beep” stimuli). Following presentation of the video, a response screen was presented at which time subjects could make a response. The inter trial interval (ITI) between trials was randomly jittered from 500 to 1500 ms.

The pre-training, post-training, and follow-up SJ assessments used to derive the TBW for both “flash-beep” and audiovisual speech stimuli each consisted of 300 trials (15 SOAs x 20 trials per SOA).

### *Perceptual Training Sessions*

Participants in either the “flash-beep” training group or audiovisual speech training group were asked to make the same judgments as the initial SJ assessment without feedback, however, trial-by-trial feedback was presented following each response which indicated informed the participants of the accuracy of their responses. The stimuli were identical to the task without feedback except that stimuli were presented at SOAs of  $\pm 150$ ,  $\pm 100$ ,  $\pm 50$  and 0 ms. Additionally, the number of trials was not equally distributed amongst each SOA but rather the true synchrony (0 ms) was presented at a 6:1 ration to asynchronous trials such that the total number of synchronous and asynchronous trials was equal. Previous studies have employed an equal number of synchronous and asynchronous trials to mitigate concerns of a potential response bias. Each training session consisted of 480 total trials (0 ms, 240 trials;  $\pm 150$ ,  $\pm 100$ ,  $\pm 50$  ms, 40 trials each). Visual feedback in the form of either a blue-green check mark (following

correct responses) or a red X (following incorrect responses) was presented following a participants response.

Participants in the audiovisual speech exposure group were exposed to the same stimuli as those subjects in the audiovisual speech training group but were asked to complete a visual detection task instead of making a simultaneity judgment. Subjects were instructed to respond “if the video is red” or “if the video is NOT red” by pressing 1 and 2 respectively on the keyboard. Additionally, each subject was shown an example of a frame from the “red” and “NOT red” videos in order to ensure that subjects understood what represented the corresponding color of the video. The red tinted video was presented on 10% of trials (90% of trials presented with normal color). The audiovisual speech exposure task is similar to previously employed tasks for low-level “flash-beep” stimuli.

### *Protocol*

Participants completed four consecutive days of perceptual training sessions and a post-training follow-up assessment that occurred 1 week after the last training session to determine the stability of training related changes in temporal acuity. Prior to the first session, participants completed a pre-training SJ assessment for the stimulus type for which the participant was not completing perceptual training (either “flash-beep” or audiovisual speech). For perceptual training days 1-3, participants completed a pre-training SJ assessment, a training session, and a post-training assessment for the stimulus type and protocol to which they were randomly assigned (i.e. low-level “flash-beep” training, audiovisual speech training, or audiovisual speech exposure). For the fourth training day, participants completed an additional training session and post-training assessment as well as a post-training SJ assessment of the non-trained stimulus. One week following the last training session (which we will refer to as the 1 week follow-up session), participants completed SJ assessment for each stimulus; no additional training was conducted on this day.

## Analysis

To estimate the TBW for each SJ assessment, each participant's individual data was fit using two sigmoid curves using the *glmfit* function in MATLAB in order to separately fit the data to measure the auditory-leading and visual-leading TBWs (the whole TBW was also measured as the sum of the auditory-leading and visual leading TBWs). To measure the TBW, we determined the point at which each best-fit sigmoid equaled a criterion value of 75% report of synchrony on the *y*-axis to determine the auditory- or visual-leading extent of the TBW which was subsequently measured relative to the participant's PSS, the point at which the sigmoids fits intersected. Group estimates of the TBW were taken as the mean of estimates that were individually derived for each participant.

## Results

Although we obtained estimates of the TBW across multiple time points during the course of perceptual training, we chose *a priori* to restrict our comparisons for statistical analysis to the baseline estimate of the TBW conducted prior to the first training session, the estimate of the TBW immediately following the last training session, and the estimate of the TBW obtained 1 week following the last training session. By limiting our estimates to these time points, we were able to analyze the most meaningful comparisons and easily compare across stimulus types for generalization of training for the untrained stimulus type. (Please refer to the Supplementary Materials, Figures 5-2 and 5-3, for estimates of the TBW obtained across all training sessions.)

Additionally, we also sought to ensure that the TBW is stable in the absence of training by having individuals complete a same SJ task as those participants completing in the audiovisual speech training group, but without feedback during perceptual training sessions as well as a pre- and post-training SJ assessment to estimate the TBW. Participants completed

four consecutive days of these practice sessions (i.e. the perceptual training SJ task without feedback). A within-subjects repeated-measures ANOVA for SJ assessments did not reveal any significant change in the whole TBW ( $F_{7,35} = 1.125, p = .370$ ), auditory-leading TBW ( $F_{7,35} = 1.329, p = .266$ ), or visual-leading TBW ( $F_{7,35} = 0.710, p = .664$ ). Practice does not appear to affect the size of the TBW in the absence of perceptual training with feedback (Figure 5-4).

### *Perceptual Training Enhances Temporal Acuity for “Flash-beep” and Audiovisual Speech Stimuli*

We first sought to determine if temporal acuity was enhanced following four consecutive days of perceptual training sessions for the two different training stimulus sets (flash-beep and speech). We first compared the estimate of the TBW obtained prior to the first training session to the estimate of the TBW obtained following the last perceptual training session. Not surprisingly and concordant with prior work (Powers et al., 2009; Powers III et al., 2016), for the “flash-beep” training group we observed an enhancement of temporal acuity following the last training session. The narrowing of the TBW was seen for the full TBW as well as for both the auditory- and visual-leading sides of the distribution (Figure 5-5A; whole TBW, from 359 ms to 201 ms,  $t_{(7)} = 4.298, p = .004$ ; auditory-leading TBW, from 143 to 91 ms,  $t_{(7)} = 3.013, p = 0.020$ ; and visual-leading group, from 216 ms to 110 ms,  $t_{(7)} = 4.629, p = .002$ ). In a similar fashion, for the audiovisual speech training group, perceptual training significantly enhanced temporal acuity as measured immediately following the last training session (Figure 5-5B; whole TBW, from 343 ms to 250 ms,  $t_{(10)} = 2.085, p = .012$ ; auditory-leading TBW, from 134 ms to 85 ms,  $t_{(10)} = 2.852, p = .017$ ; and visual-leading TBW, from 209 ms to 166 ms,  $t_{(10)} = 2.848, p = .017$ ). In contrast, we did not observe any changes in temporal acuity for those individuals merely exposed to the speech stimuli (Figure 5-5C; whole TBW from 424 ms to 446 ms,  $t_{(8)} = -0.921, p = .384$ ; auditory-leading TBW, from 191 ms to 204 ms,  $t_{(8)} = -0.638, p = .541$ ; and visual-leading TBW, from 233 ms to 243 ms,  $t_{(8)} = -0.890, p = .399$ ).

*Changes in Temporal Acuity Remain Stable for “Flash-beep” Stimuli but not Audiovisual Speech Stimuli at Least 1 Week Following Perceptual Training*

As previous studies have demonstrated that perceptual training for low-level “flash-beep” stimuli remain stable at least one week following perceptual training, we sought to replicate this finding for the “flash-beep” stimuli and determine if these changes would remain stable for audiovisual speech stimuli as well. To do so, we compared the estimate of the TBW prior to perceptual training to the estimate of the TBW measured one week following the last training session. For the “flash-beep” training group (Figure 5-5A), we observed that the whole TBW remained significantly narrower compared to the whole TBW prior to training (359 ms pre-training compared to 256 ms at follow-up,  $t_{(7)} = 2.739$ ,  $p = .029$ ). This improvement in temporal acuity was largely a result of the stability of the visual-leading TBW (216 ms pre-training compared to 132 ms at follow-up,  $t_{(7)} = 3.246$ ,  $p = 0.14$ ), as the auditory-leading TBW returned to baseline values (143 ms pre-training compared to 124 ms at follow-up,  $t_{(7)} = 0.827$ ,  $p = .435$ ). Unlike the “flash-beep” training group, the changes in temporal acuity for the audiovisual speech training group observed immediately following perceptual training (Figure 5-5B) did not exhibit the same level of stability one week following training (whole TBW, 343 ms pre-training to 278 ms at follow-up,  $t_{(10)} = 2.057$ ,  $p = .067$ ; auditory-leading TBW, from 134 ms pre-training to 96 ms at follow-up,  $t_{(10)} = 2.143$ ,  $p = .058$ ; and visual-leading TBW, from 209 ms pre-training to 182 ms at follow-up,  $t_{(10)} = 1.676$ ,  $p = .125$ ). For the audiovisual speech exposure group (Figure 5-5C), we did not observe any further changes in the temporal acuity for audiovisual speech relative to the pre-training TBW (whole TBW, from 424 ms pre-training to 491 ms at follow-up,  $t_{(8)} = -1.632$ ,  $p = .141$ ; auditory-leading TBW, from 191 ms pre-training to 228 ms at follow-up,  $t_{(8)} = -1.738$ ,  $p = .120$ ; and visual-leading TBW, from 233 ms pre-training to 263 ms at follow-up,  $t_{(8)} = -1.468$ ,  $p = .180$ ).

### *Perceptual Training does not Generalize Across Levels of Stimulus Complexity*

We sought to determine if perceptual training that enhanced temporal acuity for a simple, low-level “flash-beep” stimulus or complex, audiovisual speech stimulus generalized across stimulus complexity by first comparing the TBWs for audiovisual speech stimuli and “flash-beep” stimuli respectively prior to and immediately following perceptual training. For the “flash-beep” training group (Figure 5-6A), we did not observe a change in the temporal acuity for audiovisual speech stimuli (whole TBW, from 308 ms to 324 ms,  $t_{(7)} = -0.486$ ,  $p = .642$ ; auditory-leading TBW, from 146 ms to 157 ms,  $t_{(7)} = -0.781$ ,  $p = .0460$ ; and visual-leading TBW, from 162 ms to 167 ms,  $t_{(7)} = -0.242$ ,  $p = .815$ ). Likewise, for the audiovisual speech training group (Figure 5-6B), we did not observe a change in the temporal acuity for the “flash-beep” stimulus (whole TBW, from 336 ms to 359 ms,  $t_{(10)} = -0.654$ ,  $p = .528$ ; auditory-leading TBW, from 140 ms to 144 ms,  $t_{(10)} = -0.230$ ,  $p = .823$ ; and visual-leading TBW, from 196 ms to 215 ms,  $t_{(10)} = -0.654$ ,  $p = .528$ ).

### **Discussion**

In this study, we show that perceptual training is capable of enhancing temporal acuity for both simple, low-level (flashes and beeps) stimuli as well as for more complex stimuli (audiovisual speech). Interestingly, training on one level of stimulus complexity does not appear to generalize across levels, suggesting that the changes in temporal acuity we observe may be more stimulus specific in nature rather than an overall enhancement of audiovisual temporal acuity. Finally, we report that changes in temporal acuity for audiovisual speech stimuli following perceptual training are less durable relative to the changes in temporal acuity for simple stimuli one week following perceptual training.

Visual speech cues enhance the comprehension of speech, even if the auditory signal is clear (Arnold & Hill, 2001). These behavioral benefits appear to be related to the temporal processing of these audiovisual cues, as temporal acuity appears to correlate with speech

intelligibility (Conrey & Pisoni, 2006) and greater integration of audiovisual speech (Ryan A. Stevenson et al., 2012). Our finding that perceptual training may enhance temporal acuity for audiovisual speech stimuli is of particular importance as the temporal integration of audiovisual speech has been reported to be impaired for individuals with ASD (Mongillo et al., 2008; Ryan A. Stevenson, Siemann, Woynaroski, et al., 2014). Recently, *Stevenson et al. 2014* demonstrated that children with ASD exhibit wider TBWs than typically developed peers for audiovisual stimuli, with the major changes being seen for audiovisual speech stimuli. Furthermore, *Stevenson et al. 2014* reported that temporal acuity was strongly correlated with the magnitude of audiovisual integration (as measured by perception of the McGurk illusion) for individuals with ASD (Ryan A. Stevenson, Siemann, Schneider, et al., 2014). More specifically, those individuals with ASD with the largest TBW were those in which reports of the fused McGurk percept were lowest. These observations suggest that when audiovisual temporal acuity is impaired, cues that should be perceived separately may instead be improperly perceptually bound, which further decreases the reliability of the sensory signals and results in less efficient and weaker perceptual integration. Based on these findings, perceptual training to enhance audiovisual temporal acuity has been suggested as a possible intervention for clinical conditions for which temporal acuity is impaired (Wallace & Stevenson, 2014). Here, we observe that perceptual training with audiovisual speech, similar to that previously reported to elicit changes in temporal acuity for low-level stimuli, enhances temporal acuity. This supports the feasibility of perceptual training as a possible means of intervention to enhance temporal acuity and multisensory integration for those for whom multisensory processing is impaired. Furthermore, as training with low-level “flash-beep” stimuli does not appear to produce a generalized enhancement of temporal acuity that generalizes to speech stimuli, this suggests that future adaptations of perceptual training directed at enhancing temporal acuity for audiovisual speech in clinical conditions should employ speech stimuli during perceptual training.

Unlike the TBW for “flash-beep” stimuli, the TBW for audiovisual speech was not statistically narrower one week after the last training session (whole TBW  $p = 0.067$ ). One possible explanation as to why the effects of perceptual training may be less durable relative to the “flash-beep” stimulus is that a wider TBW for this stimulus type may be better suited to perceptually bind speech stimuli and thus the enhanced temporal acuity following perceptual training is not well suited to this stimulus. Thus, for a typically developed population, perceptual training that narrows the TBW for speech stimuli may not be reflective of the adaptive nature of an otherwise wider TBW for speech stimuli. Interestingly, prior studies have often observed that the TBW for speech stimuli is wider than simple, static stimuli (i.e. flashes and beeps) (R. A. Stevenson & Wallace, 2013) and complex non-speech stimuli (Dixon & Spitz, 1980). For speech stimuli, wider TBWs are observed for congruent audiovisual speech signals (i.e. the correct visual signal is paired with the corresponding auditory signal) compared to TBWs for artificial, incongruent speech stimuli (van Wassenhove et al., 2007). Furthermore, in the context of congruent speech stimuli, language experience influences the temporal perception of speech as individuals are reported to perceive speech in their native language (or a second language) to be most synchronous (i.e. the SOA of the PSS) with greater visual-lead relative to speech stimuli in a non-native language or language with which they are not proficient (Navarra, Alsius, Velasco, Soto-Faraco, & Spence, 2010). Collectively, these findings may suggest that the predictive nature of the visual signal favors a wider TBW than for non-speech stimuli that reflects the variable nature of the relationship of the visual and auditory cues. As speech is truly an asynchronous stimulus as the onset of visual articulatory gestures of speech actually precede the auditory speech signal by approximately 100 to 300 ms (Chandrasekaran et al., 2009), a wider but not too wide TBW may more accurately the statistics of the speech stimulus. Yet a TBW that is too wide, as is observed for individuals with ASD, may be maladaptive as auditory and visual speech cues that are not related may be perceptually bound (Ryan A Stevenson, Siemann, Schneider, et al., 2014).

Similarly, in the one week follow-up period the relative contribution of experience with speech stimuli is much greater than it is for the more artificial “flash-beep” type stimuli. Following exposure to asynchronous speech, studies have observed recalibration of the perceptual representations of the temporal relationships for speech (Van der Burg & Goodbourn, 2015; Vatakis, Navarra, Soto-Faraco, & Spence, 2007). In a similar sense, the relative asynchrony of the visual and auditory cues for natural speech may actually result in adaptation of the TBW following perceptual training. If asynchronous speech cues exceeded the range of the TBW following perceptual training, we might hypothesize that this relatively greater asynchronous signal might drive recalibration of the TBW to be wider following a week after training. Alternatively, there may exist a strong prior representation of the TBW for speech stimuli that is less susceptible to change following perceptual training. Future investigations will be necessary to determine if additional perceptual training sessions might elicit more durable changes in temporal acuity. Similarly, additional investigations will be necessary to determine if the TBW for audiovisual speech approaches the initial TBW estimate at longer than one week following perceptual training.

Our finding that enhancements in temporal acuity did not generalize across stimulus complexity suggests that perceptual training may not elicit more general changes in temporal acuity but rather stimulus specific changes. Some have posited the existence of a global mechanism for temporal acuity, given that stimulus-specific differences in the size of the TBWs are highly correlated within individuals (R. A. Stevenson & Wallace, 2013). Although reports of generalized changes in temporal acuity are limited, *Navarra et al. 2005* observed that an increase in the TBW for simple stimuli following exposure to asynchronous complex stimuli (including speech) (*Navarra et al., 2005*), suggesting that adaptation may drive general changes in temporal acuity independent of stimulus complexity. Yet unlike the study by *Navarra et al. 2005*, the current study employed active perceptual feedback and may differ in mechanism from changes in temporal acuity observed following passive adaptation. Our findings suggest that, at

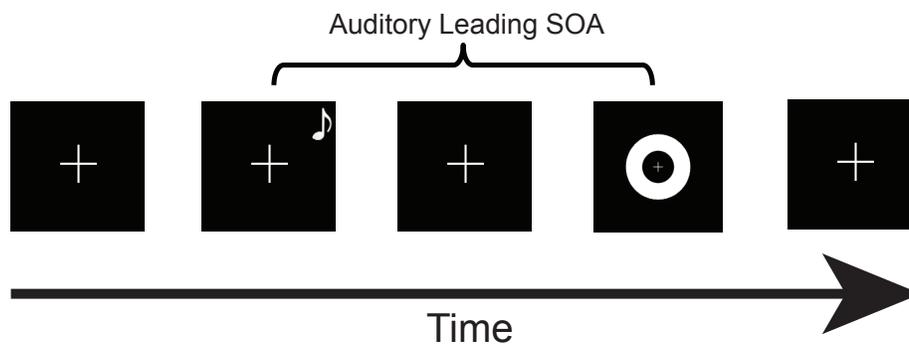
least following perceptual training, enhancements of temporal acuity do not manifest as enhancements in a global change in temporal acuity. Although not tested in the current study, it remains possible that the observed changes in temporal acuity might also be observed for unimodal stimuli with similar stimulus complexity. Recent accounts of unisensory perceptual learning observe that improvements in temporal perception transfer across sensory modalities, from the auditory modality to the visual modality, but this transfer may be task specific and has only been observed for one level of stimulus complexity (simple stimuli) (Barakat, Seitz, & Shams, 2015; McGovern, Astle, Clavin, & Newell, 2016). Future studies will be necessary to determine if the changes in temporal acuity generalize within a level of multisensory stimulus complexity (i.e. different types of speech stimuli) as temporal acuity for different types of speech stimuli themselves have been reported to vary (Vatakis & Spence, 2006b).

Collectively, our findings suggest that perceptual training is capable of enhancing temporal acuity for low-level “flash-beep” stimuli and more complex audiovisual speech stimuli. We postulate that perceptual training may have the capacity to enhance temporal acuity for individuals with impaired temporal acuity for audiovisual speech.

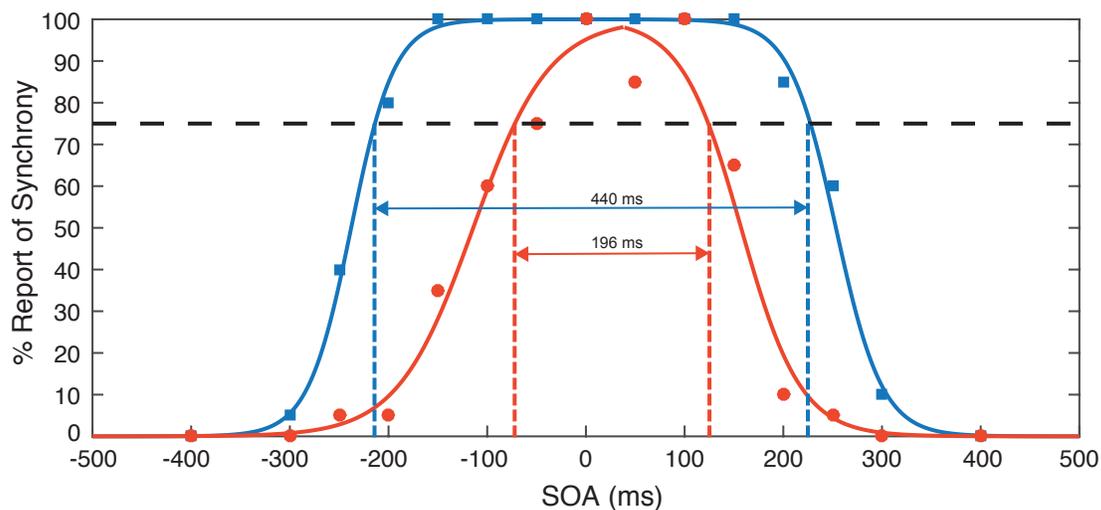
A



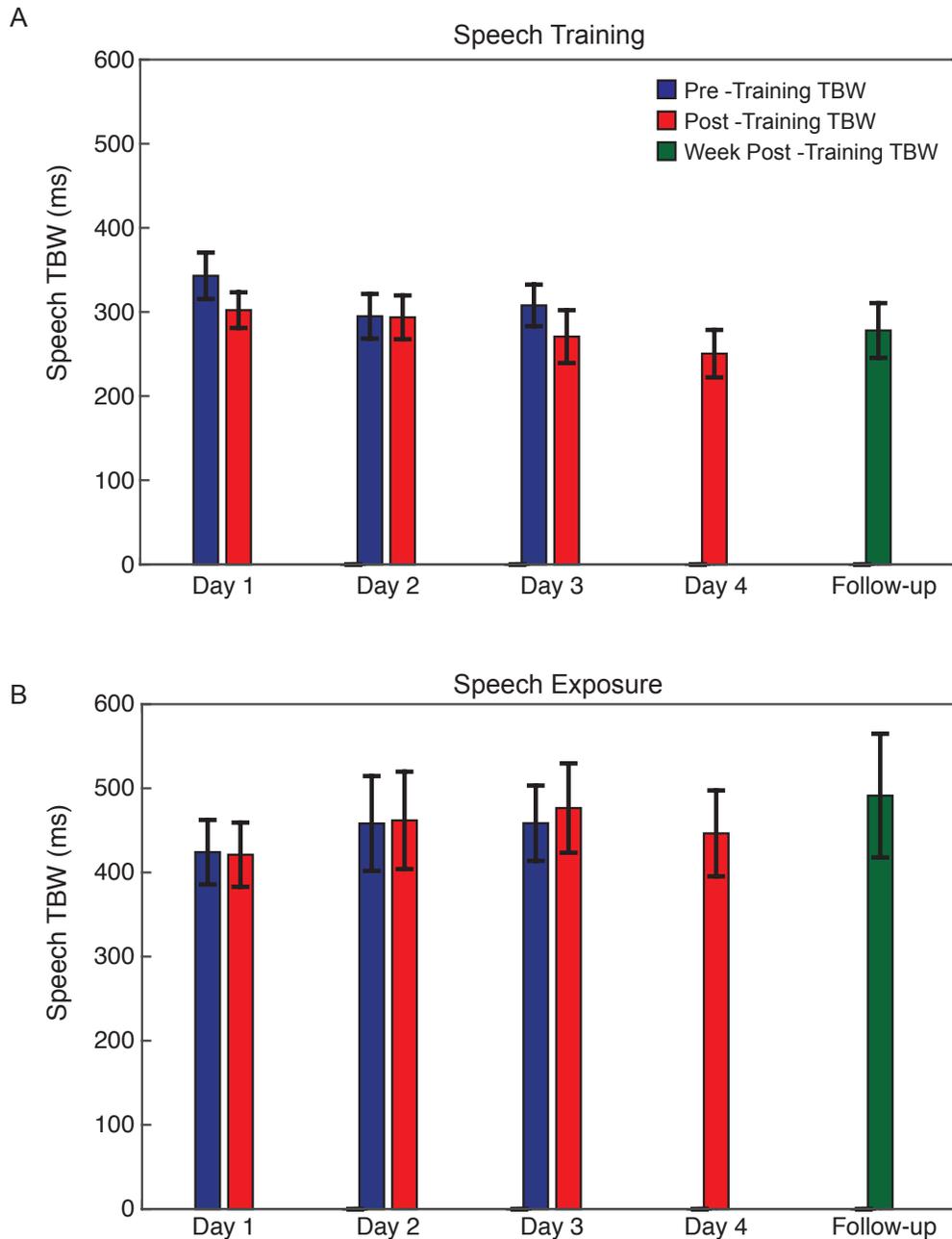
B



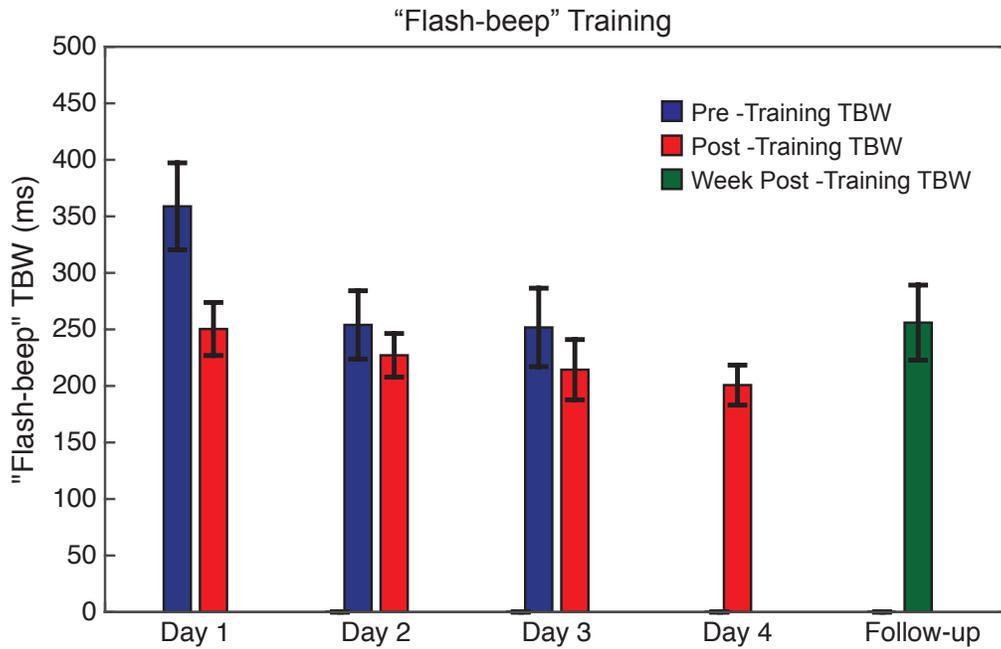
C



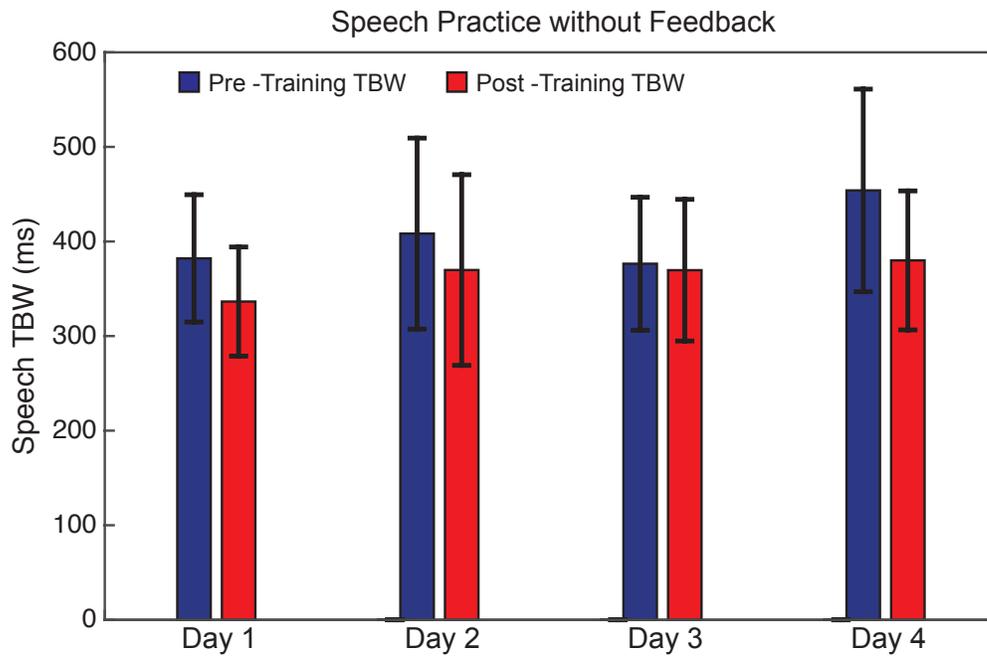
**Figure 5-1. Study Methods** A) Individual frames from the audiovisual speech stimulus video and auditory waveform. B) “Flash-beep” stimulus with an auditory leading SOA. C) Data from a representative study participant in the training group prior to perceptual training (blue squares) and following the last day of perceptual training (red squares). Sigmoid curves were fit to this participant’s data prior to perceptual training (blue lines) and following the last day of perceptual training (red lines). The whole, auditory-leading, and visual-leading TBWs were estimated by determining the epoch of time at which subject’s reported greater than 75% of trials to be synchronous.



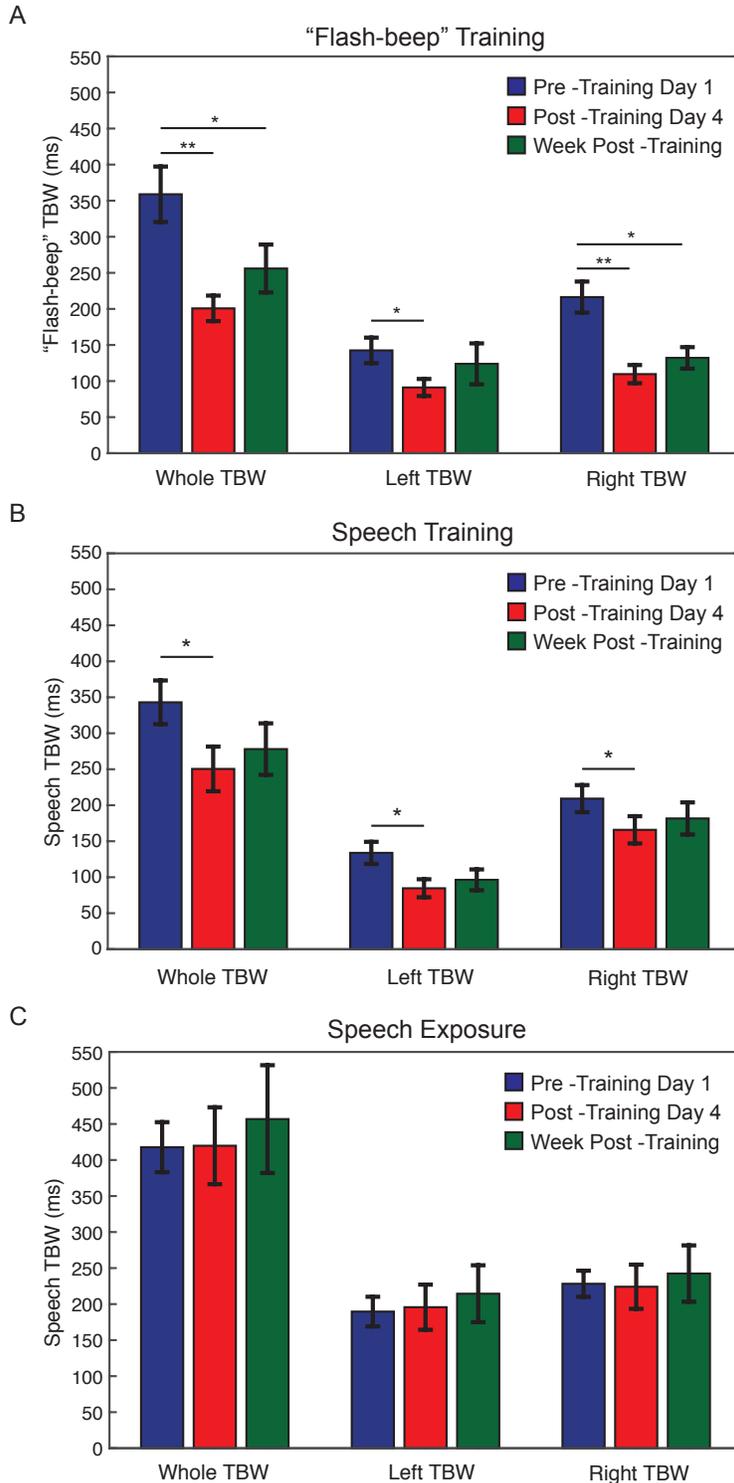
**Figure 5-2. Perceptual Training Influences Temporal Acuity for Audiovisual Speech** A) A within-subjects repeated measures ANOVA determined there was a significant main effect of SJ assessment across all pre-training, post-training, and one week following the last perceptual training session SJ assessments for the audiovisual speech training group ( $F_{7,70} = 3.116, p = .006$ ). B) A within-subjects repeated measures ANOVA determined there was a not a significant main effect of SJ assessment across all pre-training, post-training, and one week following the last perceptual training session SJ assessments for the audiovisual speech exposure group ( $F_{3,351, 26.807} = 1.006, p = .412$ ). Degrees of freedom were adjusted using a Greenhouse-Geisser correction to account for a significant violation of the assumption of sphericity.



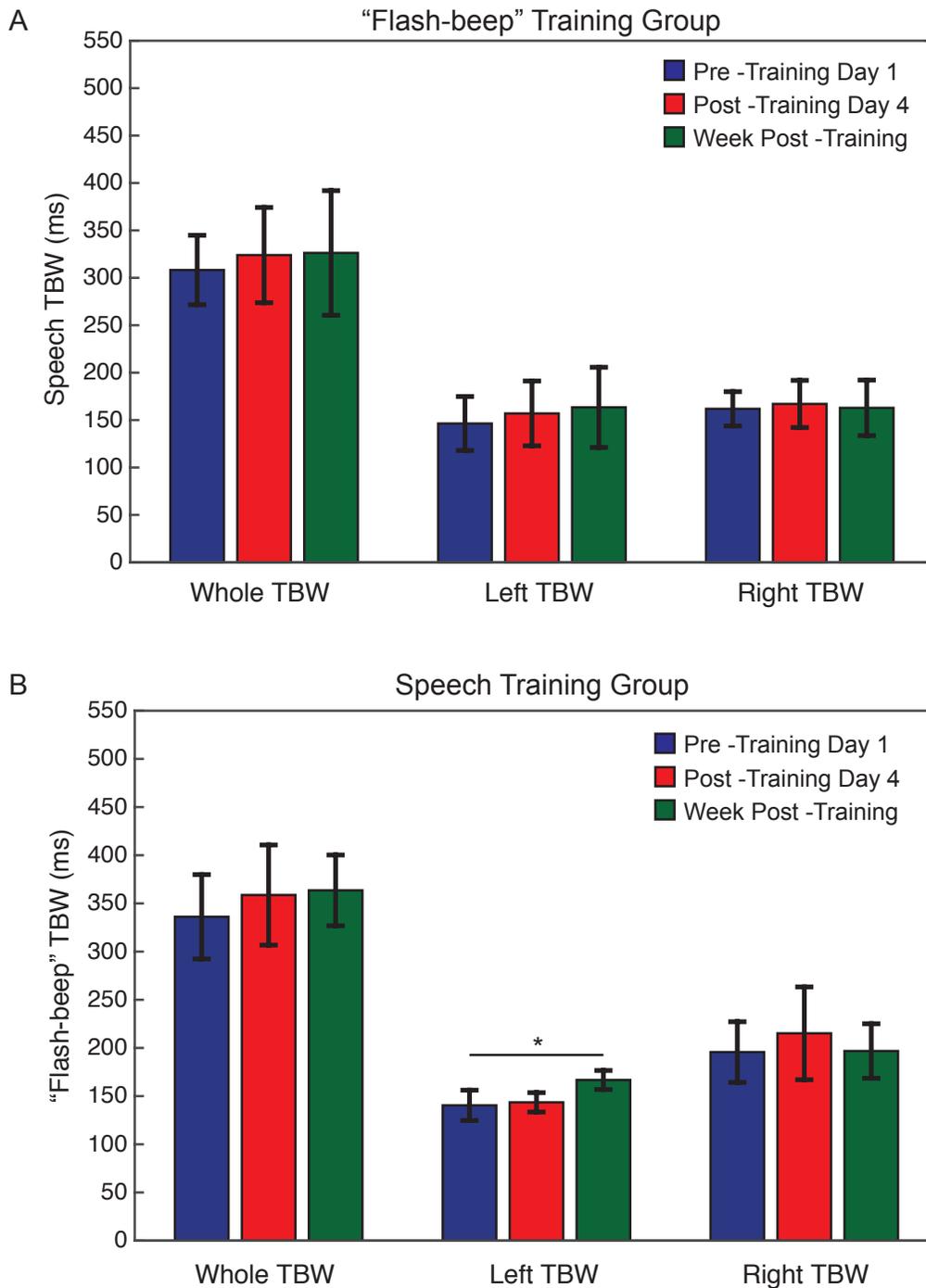
**Figure 5-3. Perceptual Training Influences Temporal Acuity for “Flash-beep” Stimuli A** within-subjects repeated measures ANOVA determined there was a significant main effect of SJ assessment across all pre-training, post-training, and one week following the last perceptual training session SJ assessments for the “flash-beep” training group ( $F_{7,49} = 7.91, p < .001$ ).



**Figure 5-4. Practice Without Feedback does not Affect Temporal Acuity for Audiovisual Speech**  
 We observe that the TBW for audiovisual speech remains stable over the course of four days for subjects who completed pre- post-training SJ assessments on each day prior to a following completion of the same SJ task as the audiovisual speech training group but without feedback.



**Figure 5-5. Perceptual Training Enhances Temporal Acuity** A) The whole TBW for the “flash-beep” training group was narrower both immediately following the last training as well as one week following the last training session. B) The whole TBW for the audiovisual speech training group was narrower immediately following the last training session while the TBW at one week following the last training session, although narrower than the pre-training TBW, was not statistically different from the TBW estimated prior to training. C) No change in the whole TBW was observed immediately following the last training session and at the one week follow-up session for the audiovisual speech exposure group.



**Figure 5-6. Perceptual Training Does Not Generalize Across Stimulus Complexity** A) No change in the TBW for audiovisual speech stimuli was observed for the “flash-beep” training group. B) No change in the TBW for “flash-beep” stimuli was observed for the audiovisual speech training group.

## References

- Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339-355.
- Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, *25*(2), R60-R61.  
doi:10.1016/j.cub.2014.12.011
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433-436.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput Biol*, *5*(7), e1000436.
- Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The journal of the acoustical society of america*, *119*(6), 4065-4073.
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388-1404.
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719-721.
- Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, *124*, 876-886.
- Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The journal of the acoustical society of america*, *108*(3), 1197-1208.
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision research*, *37*(15), 2133-2141.
- Jaekl, P., Pesquita, A., Alsius, A., Munhall, K., & Soto-Faraco, S. (2015). The contribution of dynamic visual cues to audiovisual speech perception. *Neuropsychologia*, *75*, 402-410.

- Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive brain research*, *17*(2), 447-453.
- MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British journal of audiology*, *21*(2), 131-141.
- McGovern, D. P., Astle, A. T., Clavin, S. L., & Newell, F. N. (2016). Task-specific transfer of perceptual learning across sensory modalities. *Current Biology*, *26*(1), R20-R21.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, *7*(10), 3215-3229.
- Mongillo, E. A., Irwin, J. R., Whalen, D., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of autism and developmental disorders*, *38*(7), 1349-1358.
- Navarra, J., Alsius, A., Velasco, I., Soto-Faraco, S., & Spence, C. (2010). Perception of audiovisual speech synchrony for native and non-native language. *Brain Research*, *1323*, 84-93.
- 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. *Cognitive brain research*, *25*(2), 499-507.
- Olasagasti, I., Bouton, S., & Giraud, A.-L. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex*.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*(4), 437-442.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, *32*(18), 6263-6274.

- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, *29*(39), 12265-12274.
- Powers III, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports*, *6*, 23374.
- Quinto, L., Thompson, W. F., Russo, F. A., & Trehub, S. E. (2010). A comparison of the McGurk effect for spoken and sung syllables. *Attention, Perception, & Psychophysics*, *72*(6), 1450-1454.
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147-1153.
- Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), B69-B78.
- Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current opinion in neurobiology*, *17*(2), 148-153.
- Seitz, A. R., Nanez, J. E., Sr., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, *6*(9), 966-973. doi:10.1167/6.9.9
- Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia*, *61*, 259-268.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, *12*(11), 411-417.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255-266.

- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*, *34*(3), 691-697.
- Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Brief report: arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of autism and developmental disorders*, *44*(6), 1470-1477.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249-261. doi:10.1007/s00221-013-3507-3
- Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research*, *225*(4), 479-489. doi:10.1007/s00221-012-3387-y
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. *Journal of Experimental Psychology-Human Perception and Performance*, *38*(6), 1517-1529. doi:10.1037/a0027339
- Sumbly, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26*(2), 212-215.
- Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *335*(1273), 71-78.
- Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B-Biological Sciences*, *282*(1804). doi:10.1098/rspb.2014.3083

- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181-1186.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598-607.
- Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, *181*(1), 173-181.
- Vatakis, A., & Spence, C. (2006a). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, *1111*(1), 134-142.
- Vatakis, A., & Spence, C. (2006b). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience letters*, *393*(1), 40-44.
- Venezia, J. H., Fillmore, P., Matchin, W., Isenberg, A. L., Hickok, G., & Fridriksson, J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *NeuroImage*, *126*, 196-207.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871-884.
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105-123.
- Zweig, L. J., Suzuki, S., & Grabowecky, M. (2015). Learned face-voice pairings facilitate visual search. *Psychonomic Bulletin & Review*, *22*(2), 429-436. doi:10.3758/s13423-014-0685-

## Chapter VI

### The Neural Correlates of Enhanced Temporal Acuity for Audiovisual Speech Following Multisensory Perceptual Training

#### Abstract

Auditory and visual speech cues that occur proximately in time are more likely to be integrated and subsequently influence the perception speech, often manifesting in enhanced intelligibility and comprehension. Temporal acuity for audiovisual speech may be illustrated by the construct of a temporal binding window (TBW), which describes the epoch of time during which objectively asynchronous stimuli may be perceptually bound. Recently, numerous studies have reported that perceptual training is capable of enhancing multisensory, particularly for audiovisual speech. As better multisensory temporal acuity (i.e. smaller TBWs) have been reported to be correlated with greater audiovisual integration, it has been hypothesized that perceptual training to enhance multisensory temporal acuity may enhance the benefits derived from the integration of audiovisual speech cues. In this study, we sought to determine the neural substrates of the changes in temporal acuity for audiovisual speech that occur following multisensory perceptual training using fMRI. We report that perceptual training appears to drive significant changes in cortical activity in both sensory and decisional areas. These changes are much more extensive than the neural changes following perceptual training with simple “flash-beep” stimuli, suggestive that multisensory perceptual training that enhances temporal acuity for audiovisual speech engages a wider cortical network.

#### Introduction

The integration of sensory signals from separate sensory modalities provides numerous behavioral benefits (Diederich & Colonius, 2004; Van der Burg, Cass, Olivers, Theeuwes, &

Alais, 2010), particularly in regards to speech and communication (Schwartz, Berthommier, & Savariaux, 2004; Williams, Darcy, & Newman, 2016). For speech, the benefit of multisensory processing is readily apparent as visual speech cues (with the auditory signal) facilitate of speech comprehension in noisy environments (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumbly & Pollack, 1954) or when speech is difficult to understand (Arnold & Hill, 2001). Important to whether multisensory signals will be integrated is the temporal structure of multisensory stimuli such that stimuli that are more proximate in time are more likely to be perceptually bound. The temporal binding window (TBW) has been utilized to describe the epoch of time for which temporally proximate stimuli are likely to be perceptually (Wallace & Stevenson, 2014). The TBW subsequently provides an estimate of temporal acuity for a particular stimulus. Furthermore, the size of the TBW has been reported to be correlated with the magnitude of multisensory integration (Ryan A. Stevenson, Zemtsov, & Wallace, 2012).

Perceptual representations of multisensory temporal relationships remain plastic in adulthood. Recently, numerous studies have demonstrated that perceptual training is capable of enhancing multisensory temporal acuity for simple, “flash-beep” like stimuli (Powers, Hillock, & Wallace, 2009) as well as audiovisual speech (Chapter 5). These studies have also observed that multisensory perceptual training may also elicit generalized to enhancement of unisensory (visual) temporal acuity (Powers III, Hillock-Dunn, & Wallace, 2016). At the neural level, changes in temporal acuity for “flash-beep” stimuli have been reported to arise from changes in a sensory network focused on the posterior superior temporal sulcus (pSTS) (Powers, Hevey, & Wallace, 2012). Following perceptual training to enhance temporal acuity for “flash-beep” stimuli, Powers *et al.* (2012) observed decreased activation in the pSTS, regions of visual and auditory cortex, and the superior cerebellum as well as changes in effective connectivity from a feedforward network from auditory and visual cortex to pSTS prior to perceptual training to a more distributed network involving feedback from pSTS to visual and auditory cortex after perceptual training (Powers *et al.*, 2012). Furthermore, Powers *et al.* (2012) observed changes

in resting state connectivity that reflected increased coupling of the pSTS with areas such as the superior colliculus (SC). Interestingly, the findings of Powers *et al.* (2012) appear to reflect changes in primarily sensory areas as well as areas such as the pSTS that is known to be important for multisensory processing.

Recently, perceptual training has also been reported to narrow the TBW for audiovisual speech (Chapter 5). Here we sought to determine if the changes in neural activity that develop following perceptual training to enhance temporal acuity for audiovisual speech are similar to those changes in neural activity following perceptual training to enhance temporal acuity for “flash-beep” stimuli. In comparison to “flash-beep” stimuli, audiovisual speech is a much more complex stimulus. The TBW for audiovisual speech is typically larger than for a “flash-beep” stimulus and temporal acuity for audiovisual may be influenced by various factors including semantic relationships between the auditory and visual stimuli (Vatakis & Spence, 2007). The perception of synchronous and asynchronous speech is observed across numerous neural regions beyond primary sensory areas including multisensory superior temporal cortex (Ryan A Stevenson, Altieri, Kim, Pisoni, & James, 2010; Ryan A Stevenson, VanDerKlok, Pisoni, & James, 2011). Furthermore, understanding how changes in temporal acuity develop for multisensory speech stimuli may be important as temporal acuity for audiovisual speech is suggested to be impaired in developmental disorders, such as autism spectrum disorder (ASD) (Ryan A Stevenson *et al.*, 2014), that are characterized by deficits of communication (Wallace & Stevenson, 2014).

In the current study, we sought to identify the neural changes using fMRI that develop following perceptual training to enhance multisensory temporal acuity for speech stimuli. We employed a similar multisensory training paradigm that has been utilized by previous studies (Powers *et al.*, 2009; Powers III *et al.*, 2016). Participants completed two fMRI neuroimaging session prior to and following multisensory perceptual training. We report that perceptual training enhances temporal acuity for audiovisual speech and that the changes in neural activity

following perceptual training are more widespread for audiovisual speech in comparison to the changes in neural activity following perceptual training to enhance temporal acuity for “flash-beep” stimuli.

## **Methods**

### *Subjects*

A total 10 adult individuals were recruited to participate in the study. All participants had self-reported normal hearing and normal or corrected to normal vision. All recruitment and experimental procedures were approved by the Vanderbilt University Institutional Review Board and in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### *Experimental Procedure*

The experimental sessions for this study were completed over four consecutive days. The fMRI neuroimaging protocol consisted of two sessions for which participants completed two separate MRI scans were conducted on the day prior to the first perceptual training session and on the day following the last perceptual training session. The multisensory perceptual training protocol consisted of two sessions conducted on two separate days.

## **Multisensory Perceptual Training Protocol**

### *Stimuli*

For all tasks, participants were seated in a light and sound attenuating WhisperRoom™ (SE 2000 Series, Whisper Room Inc.). The speech stimulus for all tasks consisted of a video of a female speaker uttering the phonemic unit /ba/ for both SJ assessments and perceptual training blocks. The video spanned an area of 1920 x1080 pixels and was presented in color at approximately 48 cm from the participants on a monitor (Samsung syncmaster 22 inch 2233 RZ

LCD) with a refresh-rate of 60 Hz. The total length of each video was cropped with digital video editing software (Adobe Premiere Pro) such that each clip started and ended in a neutral, mouth-closed position. The duration of the video was 1850 ms and included all pre-articulatory gestures. The auditory signal was presented binaurally via headphones (Sennheiser HD 558) with no interaural time or level differences. The intensity of all auditory stimuli was calibrated using a sound level meter (Larson Davis SoundTrack® LxT2). All SJ tasks were presented using E-Prime version 2.0.8.

### *SJ Assessments*

A pre-training and post-training SJ assessment was conducted to obtain an estimate of temporal acuity prior to and following the perceptual training session. For all SJ tasks, participants were instructed to judge whether the auditory and visual stimuli were “synchronous...properly lined up” or “asynchronous...were NOT properly lined up” by pressing 1 or 2 respectively using a keyboard. Participants were asked to maintain fixation on the fixation marker (1 cm x 1 cm) that was present during the inter trial interval (ITI) but not during the presentation of the video. For the pre- and post-training SJ assessments, the auditory and visual components were presented at true synchrony (0 ms of asynchrony) or at one of the following SOAs:  $\pm 400$ ,  $\pm 350$ ,  $\pm 300$ ,  $\pm 250$ ,  $\pm 200$ ,  $\pm 150$ ,  $\pm 100$ , and  $\pm 50$ . Each trial consisted of a single presentation of the audiovisual speech stimulus. Following presentation of the video, a response screen was presented at which time subjects could make a response. The ITI between trials was randomly jittered from 500 to 1500 ms. Each SJ assessment consisted of 340 trials (17 SOAs x 20 trials per SOA).

### *Perceptual Training*

During the multisensory perceptual training sessions, participants also completed a SJ task, however, trial-by-trial feedback was presented following each response, which indicated

informed the participants of the accuracy of their responses. The audiovisual speech stimulus was identical to the stimulus presented during the pre- and post-training SJ assessments except that stimuli were presented at SOAs of  $\pm 150$ ,  $\pm 100$ ,  $\pm 50$  and 0 ms. Additionally, the number of trials was not equally distributed amongst each SOA but rather the true synchrony (0 ms) was presented at a 6:1 ration to asynchronous trials such that the total number of synchronous and asynchronous trials was equal. Other studies that have utilized a similar perceptual training protocol have employed an equal number of synchronous and asynchronous trials to mitigate concerns of a potential response bias. Each training session consisted of 480 total trials (0 ms, 240 trials;  $\pm 150$ ,  $\pm 100$ ,  $\pm 50$  ms, 40 trials each). Visual feedback in the form of either a blue-green check mark (following correct responses) or a red X (following incorrect responses) was presented following a participants response

### *Behavioral Analysis*

For both pre-/post-training SJ and SJ made during MRI sessions, we sought to determine if the overall group probability of report of synchrony changed after perceptual training sessions utilizing the mean report of synchrony at each SOA to produce grand averages. A SOA x perceptual training status (i.e. pre- or post-training) repeated-measures ANOVA was conducted for which analysis of simple effects was conducted if a significant training status x SOA interaction existed. *Post hoc* pairwise comparisons of simple effects were conducted with Bonferroni correction to determine if perceptual training significantly altered the report of synchrony at each SOA. Degrees of freedom were adjusted for all interactions where sphericity was violated using a Greenhouse-Geisser correction.

To estimate the TBW for each pre- and post-training SJ assessment, each participant's individual data was fit with two sigmoid curves using the *glmfit* function in MATLAB. This fitting method allowed for the date to be separately fit in order to measure the auditory-leading and visual-leading TBWs (the whole TBW was also measured as the sum of the auditory-leading

and visual leading TBWs). To measure the TBW, we determined the point at which each best-fit sigmoid equaled a criterion value of 75% report of synchrony on the y-axis to determine the auditory- or visual-leading extent of the TBW which was subsequently measured relative to the participant's PSS, the point at which the sigmoids fits intersected. This data fitting method and criterion value (i.e. 75% report of synchrony) has been utilized by numerous prior studies to estimate the TBW.

### **fMRI Neuroimaging Protocol**

Each participant completed two identical neuroimaging sessions with the first neuroimaging session occurring one day before the first perceptual training session and the second neuroimaging session occurring one day after the last perceptual training session. The scanning protocol consisted of one T1-weighted anatomical scan, a blocked-design localizer scan, and four event-related fMRI scans.

#### *Block-design localizer*

A block-design localizer was used to generate the regions-of-interest (ROIs). Each block contained 10 two-second trials, one word per trial, followed by 10 seconds of fixation baseline. Each trial contained a single word from a bank of digital video recordings of 30 single-syllable words (e.g. "view", "door", "make") spoken by a female native English speaker. Auditory-only words consisted of the auditory component of each video with a white visual fixation crosshairs and visual-only words consisted of only the visual component of the video recording. The localizer scan series contained six blocks (two auditory-only, two visual-only and two audiovisual blocks in random order). Subjects were instructed to pay attention to each stimulus and that they would not need to use the response box during the run.

### *Simultaneity Judgment Task*

For the main experiment, stimuli were presented in two-second trials in a rapid event-related design. The SJ tasks utilized the same speech stimulus as the behavioral SJ tasks for the multisensory perceptual training protocol. Each stimulus consisted of a speaker uttering /ba/ and was 1850 ms in length and followed by fixation crosshairs for the remainder of the trial. After each scan series subjects were asked if they could hear the stimuli presented and if any volume adjustments were necessary. Each trial contained the syllable “ba” at one of five different SOA’s (+400, +200, 0 ms) and fixation trials (fixation crosshairs only). After each stimulus trial subjects were presented with a response screen instructing them to press “1” if they believed the visual and auditory speech information occurred at the same time and “2” if they believed the visual and auditory speech information occurred at different times. Participants were instructed prior to the beginning of the neuroimaging session what keys on the response box corresponded to a response of “1” or “2”.

Subjects viewed synchronous and asynchronous speech stimuli presented in a rapid event-related design over four runs. Each run contained 10 trials of each SOA and 35 null trials for a total of 135 dynamics.

### *MRI and fMRI Analysis*

One T1-weighted MP-RAGE anatomical MRI scans was collected at the beginning of each scanning session with a 3 Tesla whole-body MR scanner (Philips Intera Achieva, Phillips Medical Systems) with a 32-channel head coil. This scan was then used to create a cortical surface model using FreeSurfer (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999) for visualization in SUMA (Argall, Saad, & Beauchamp, 2006). For the fMRI scan series, T2\* weighed images were collected using gradient echo-planar imaging (TR = 2000 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.50 x 2.50 mm. Auditory stimuli were presented through MRI-compatible in-ear headphones (STAX SRS-005), which were covered with

earmuffs to reduce the amount of noise from the scanner. Visual stimuli were projected onto frosted glass by a projector and viewed through a mirror attached to the head coil. Responses to the target trials were collected using a five button response pad (Rowland Institute of Science, Boston, MA). Analysis of the functional scan series was conducted using Analysis of Functional NeuroImages (AFNI) (Cox, 1996).

### *Whole-brain analysis*

In order to generate whole brain maps of the amplitude and standard deviation measures at each voxel, we carried out a voxel-wise analysis using the AFNI function *3dDeconvolve*, which uses maximum-likelihood estimation in the context of the generalized linear model (GLM). For the whole-brain voxel-wise analysis, subjects' individual data were first aligned to the N27 atlas brain (Mazziotta et al., 2001) using the AFNI function *auto\_tlrc*. Blurring kernels of approximately 3-6 mm have been found to be the most sensitive for detecting activation clusters (Skudlarski, Constable, & Gore, 1999). We chose a 3 x 3 x 3 mm FWHM Gaussian kernel to minimize blurring between adjacent ROIs.

To conduct a voxel-wise search for any differences in response amplitude, the average response amplitude (beta coefficient of response relative to fixation baseline) was calculated in each voxel in each subject. *3dANOVA3* was used to perform a two-way, mixed model ANOVA with SOA and time point (pre/post) as fixed factors and subject as a random factor. After the voxel-wise ANOVA we performed a clustering technique (Xiong, Gao, Lancaster, & Fox, 1995). This finds only voxels that are significantly active above a particular threshold and spatially contiguous. The probability of finding two voxels above a particular threshold *and* being adjacent is much smaller than the chance of a single voxel above that threshold (Forman et al., 1995). Using the AFNI program *3dClustSim*, we estimated that a cluster with a size of 25 voxels would have a corrected p-value of 0.01 using first-nearest neighbor clustering. Only regions larger than 200 mm<sup>3</sup> (25 voxels) are reported.

## Results

### *Perceptual Training Enhances Temporal Acuity for Audiovisual Speech*

We sought to determine if multisensory perceptual training enhanced temporal acuity for audiovisual by determining if perceptual training affected participant's report of synchrony as well as the size of the TBW. Although we measured the TBW using a pre-training and post-training SJ assessment, we restricted our comparisons for the purpose of these analyses to comparisons between the behavioral estimates prior to the first perceptual training session and following the last perceptual training session.

We first sought to determine if an interaction of SOA and perceptual training status was present such that the perception of synchrony decreased at certain SOAs. A SOA x pre-/post-training status within-subjects repeated measures ANOVA was conducted comparing the report of synchrony at each SOA prior to perceptual training and following the last perceptual training session. A significant interaction of SOA x pre/post-training status was observed ( $F_{16,144} = 1.724$ ,  $p = .048$ ). *Post hoc* comparisons were conducted for the simple effect of pre-/post-training status determined that the probability of simultaneity judgment decreased at SOAs of -150 ms ( $p = .017$ ), -50 ms ( $p = .048$ ), 0 ms ( $p = .015$ ), 100 ms ( $p < .001$ ), and 150 ms ( $p = .011$ ).

Next, the size of the whole, auditory-leading, and visual-leading TBWs was compared prior to perceptual training and following the last perceptual training session using paired samples t-tests. The whole TBW was observed to significantly decrease from 311 ms to 241 ms ( $t_9 = 3.352$ ,  $p = .008$ ) following perceptual training. The auditory-leading TBW was also observed to significantly decrease following perceptual training from 113 ms to 79 ms ( $t_9 = 5.409$ ,  $p < .001$ ). The visual-leading TBW trended towards a decrease from 198 ms to 162 ms, however this change in the visual-leading TBW did not reach statistical significance ( $t_9 = 1.928$ ,  $p = .086$ ).

### *Report of Synchrony for fMRI SJ Assessments is Unaffected by Perceptual Training*

We also sought to determine if perceptual training altered the probability that participants reported the audiovisual speech stimuli to be synchronous at the SOAs utilized during the fMRI SJ assessments. A SOA x pre-/post-training status within-subjects repeated measures ANOVA was conducted comparing the report of synchrony at each SOA for the SJ participants during the pre-training scan to the post-training scan. A significant main effect of SOA ( $F_{1.68, 15.15} = 167.67, p < .001$ ) was observed however the main effect of pre-/post-training status ( $F_{1, 9} = 0.390, p = .548$ ) and interaction of SOA x pre/post-training status ( $F_{1.84, 16.57} = 1.570, p = .237$ ) were not significant. Thus, while multisensory perceptual training enhanced temporal acuity for audiovisual speech as measured during the post-training SJ outside of the MRI sessions, temporal acuity did not appear to be explicated changed for the asynchronous stimuli presented during the MRI sessions.

### **Whole Brain Neuroimaging Analysis**

#### *Main Effects of Training Status and SOA*

We first examined the main effect of training status for structures that showed a difference in neural activity following perceptual training (Figure 6-2, Table 6-1, 6-2, and 6-7). We observed significant clusters in the middle temporal gyrus (MTG), insula, medial frontal gyrus, and cingulate gyrus. An examination of the main effect for SOA revealed significant clusters across areas of visual and parietal cortex as well as the MTG, insula, and superior frontal gyrus.

#### *Interaction of Training Status and SOA*

We next examined areas showing a significant interaction of SOA and training status (Figure 6-3, Table 6-3, 6-4, and 6-8). While we observed a large number of clusters showing a significant interaction effect, show of the large cluster of activation were observed in the superior

frontal gyrus, inferior parietal lobule, superior colliculus, insula, middle frontal gyrus posterior cingulate and cerebellum.

### *Simple Effects of Training Status at Each SOA*

We further sought to explore whether perceptual training elicited distinct changes in neural activity at each SOA (Figure 6-4, Table 6-5, 6-6, and 6-9). Some surprisingly, we observed the greatest amount of changes in neural activity in terms of number of clusters at the auditory-leading SOAs, particular at the -400 ms SOA. A number of significant cluster were observed in frontal regions for the -400 ms SOA, including the medial frontal gyrus and inferior frontal gyrus, as well as regions of parietal cortex including the angular gyrus and supramarginal gyrus. Interestingly, at the -200 ms SOA, the majority of clusters are found in regions of visual cortex. At synchrony, significant clusters were found at the precentral gyrus and postcentral gyrus as well as some small clusters at frontal regions such as the middle frontal gyrus. The +200 ms SOA appeared to exhibit more distinct clusters with clusters in the MTG, fusiform gyrus, superior temporal gyrus (STG) as well as clusters in the insula, postcentral gyrus, and middle frontal gyrus. For the -400 ms SOA, a few clusters were observed including in the posterior cingulate and MTG.

### **Discussion**

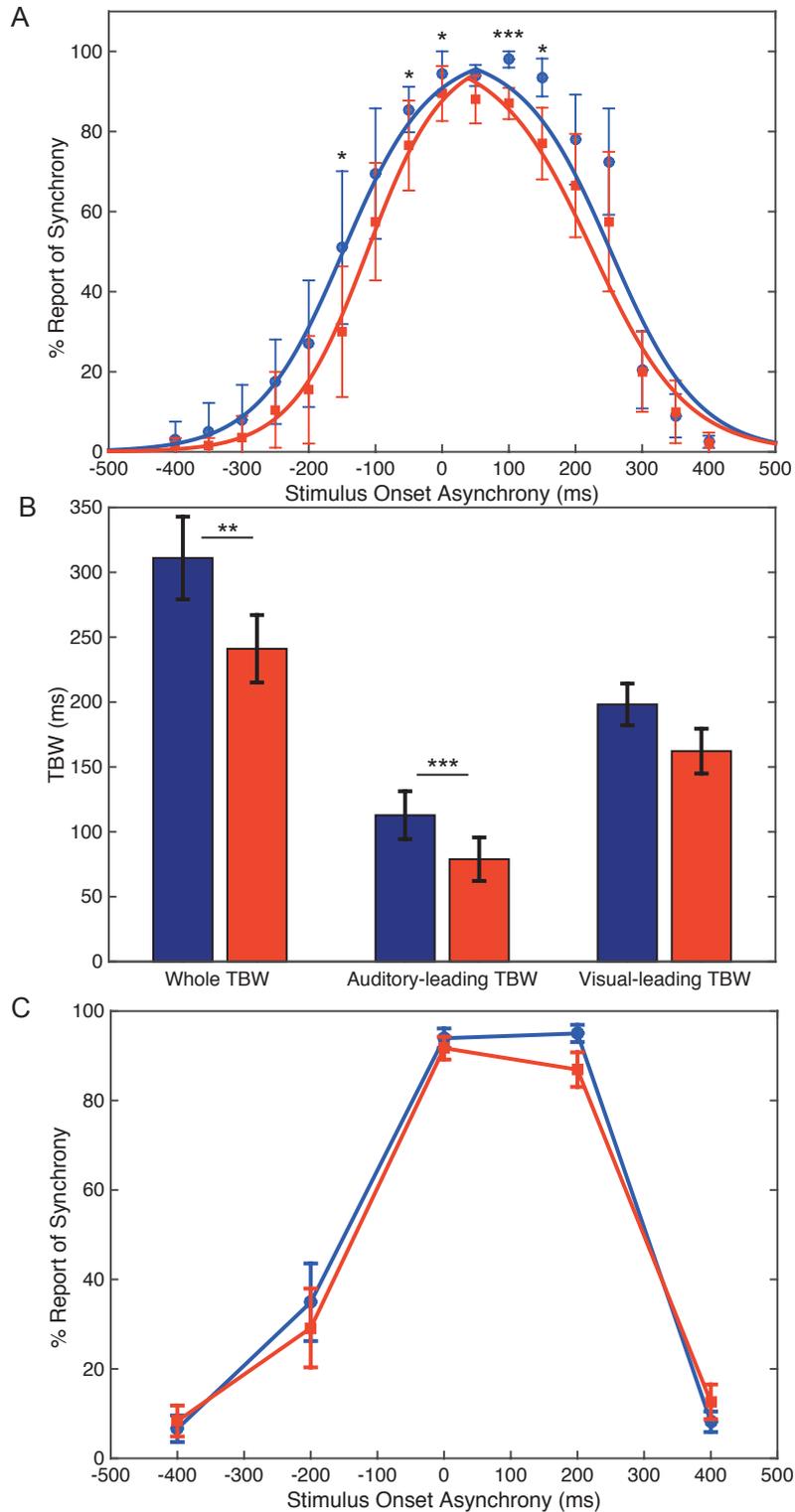
In this study we report that multisensory perceptual training to enhance temporal acuity for multisensory speech elicits more extensive changes in neural activity in comparison to the changes in neural activity observed for perceptual training to enhance temporal acuity for “flash-beep” stimuli. Although we do observe changes in activation of numerous areas that have been identified as being important for multisensory processing, we do not observe any specific changes in activity localized at the pSTS. We do however observe numerous changes in

activation in parietal and frontal regions, suggestive that changes in neural activity following perceptual training may affect perceptual decision-making.

While other forms of perceptual learning, such as visual perceptual learning were initially hypothesized to occur in sensory cortex, changes in more cognitive regions have been reported by more recent accounts of visual perceptual learning. In particular, changes in the activity of frontal cortices have been suggested to be related to changes in perceptual decision making (Kahnt, Grueschow, Speck, & Haynes, 2011). Various perceptual reweighting models of visual perceptual learning have been developed to describe changes in the readout of a decision unit as well as enhancement of the sensory information (Barbara Anne Doshier & Lu, 1999; B. A. Doshier & Lu, 1998; Petrov, Doshier, & Lu, 2005). The reweighting models suggest that visual perceptual learning occurs as a result of reweighting between the sensory representation and the decision unit (Barbara Anne Doshier & Lu, 1999; B. A. Doshier & Lu, 1998). Our findings appear to support the involvement of a more extensive cognitive regions that may be related to perceptual decision-making.

Here we also observe changes in numerous areas that have been previously reported to be important for the perception of synchrony including the SC, insula, (Calvert, Hansen, Iversen, & Brammer, 2001), STG, supplementary motor area (SMA), left post central gyrus, left (Marchant, Ruff, & Driver, 2012), and fusiform gyrus (FG) (Ryan A Stevenson et al., 2010). Interestingly, however, we do not observe some of the changes in activity in superior temporal cortex in response to synchrony and asynchrony that are observed by some studies. The difference in reported effects for activity in superior temporal cortex may be due to prior studies using responsiveness to synchrony to define ROIs (Powers et al., 2012; Ryan A Stevenson et al., 2011). The incorporation of functional localizers to define multisensory responsive and synchrony responsive ROIs in future studies of multisensory perceptual learning (for temporal acuity) may be valuable.

Our findings suggest that perceptual training to enhance temporal acuity for multisensory speech elicits changes in activity across numerous cognitive and sensory regions. Future investigations will be necessary to describe how specific changes in cortical regions parallel multisensory perceptual learning.



**Figure 6-1. Perceptual Training Enhances Temporal Acuity for Audiovisual Speech** A) A decrease in the report of synchrony was observed at SOAs of -150 ms, -50 ms, 0 ms, 100 ms, and 150 ms following perceptual training. B) A decrease in the whole and auditory-leading TBWs was observed following perceptual training. C) A significant effect of perceptual training was not observed for the SJ task during the neuroimaging session. Error bars, SEM.  $p < .05$ , \*;  $p < .01$ , \*\*;  $p < .001$ , \*\*\*.

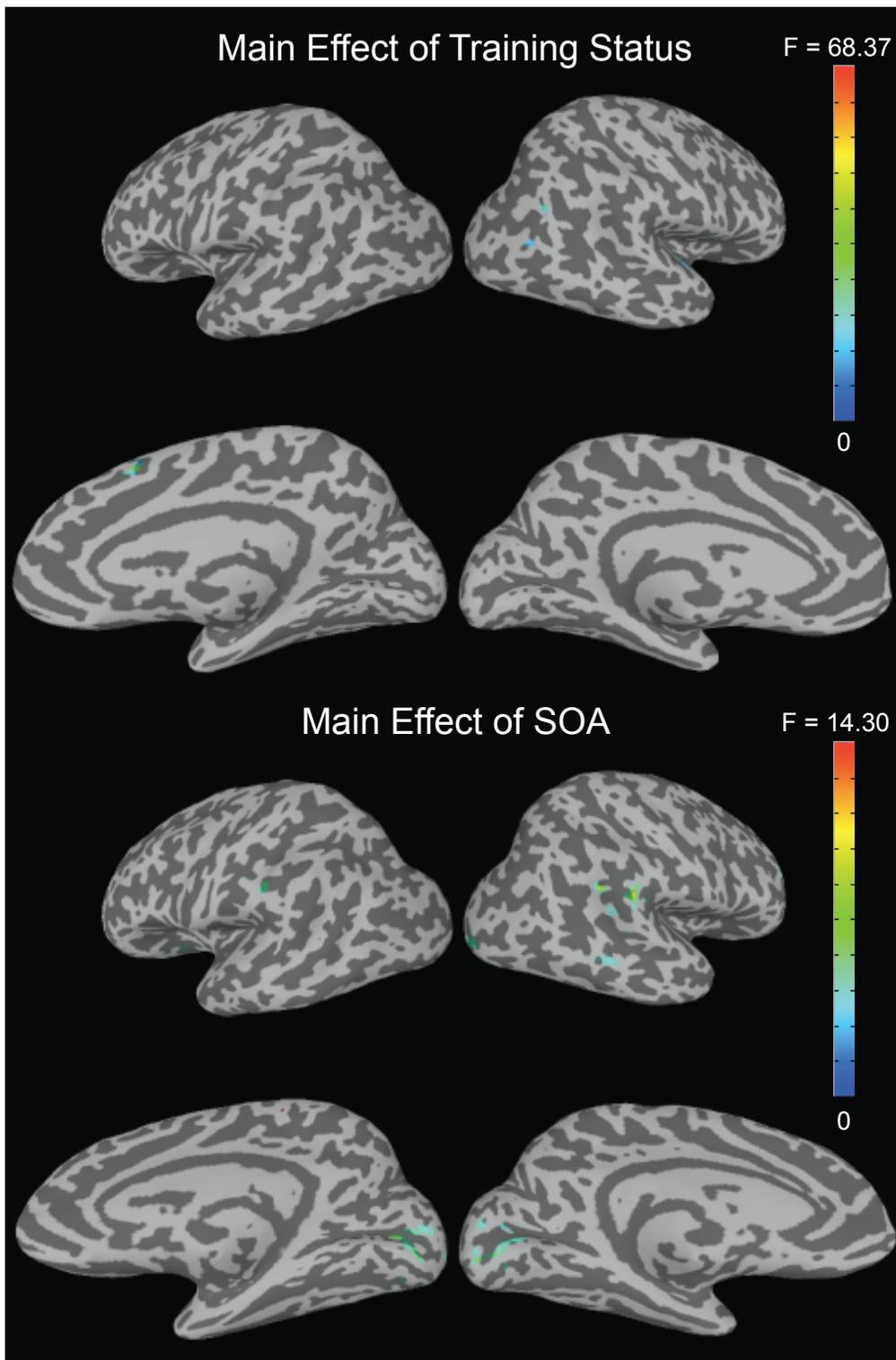


Figure 6-2. Whole brain analysis of the main effects of training status and SOA

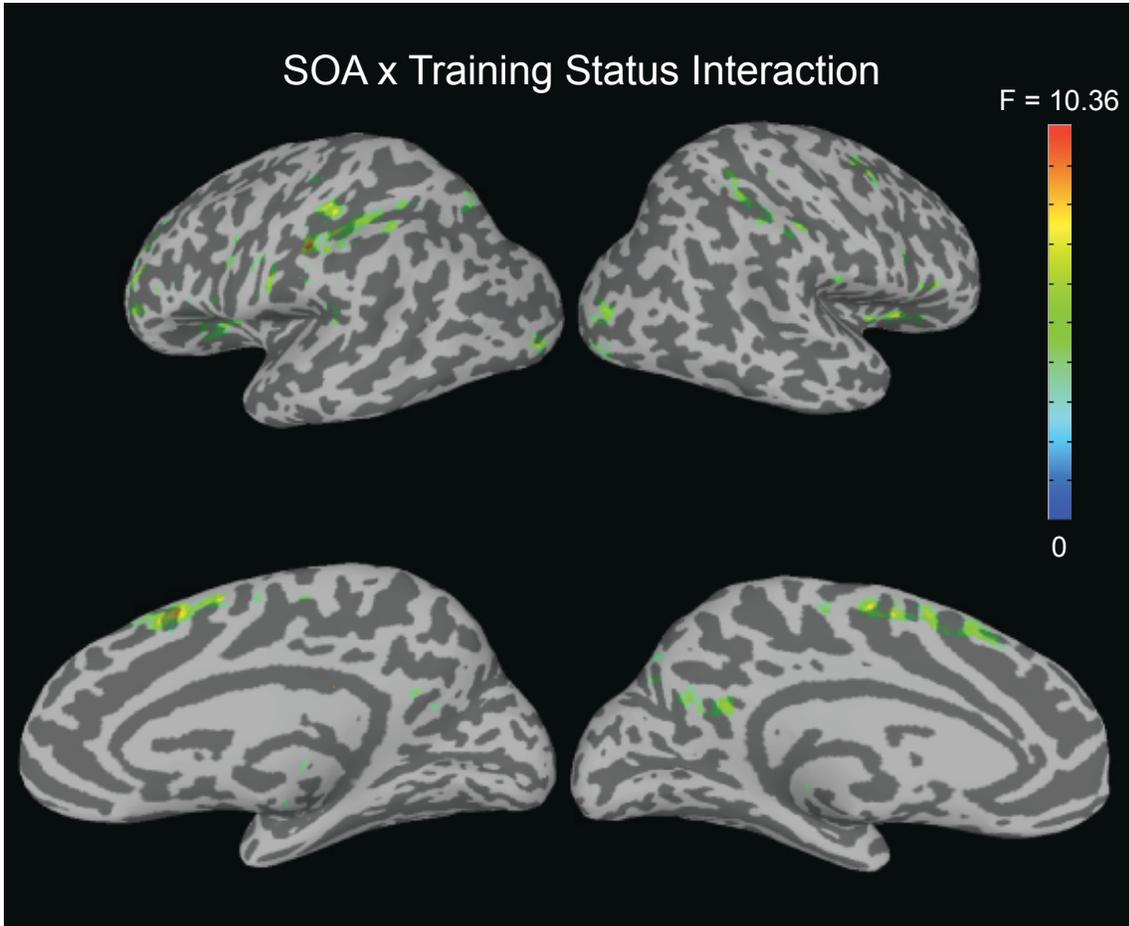


Figure 6-3. Whole brain analysis of interaction of SOA and training status

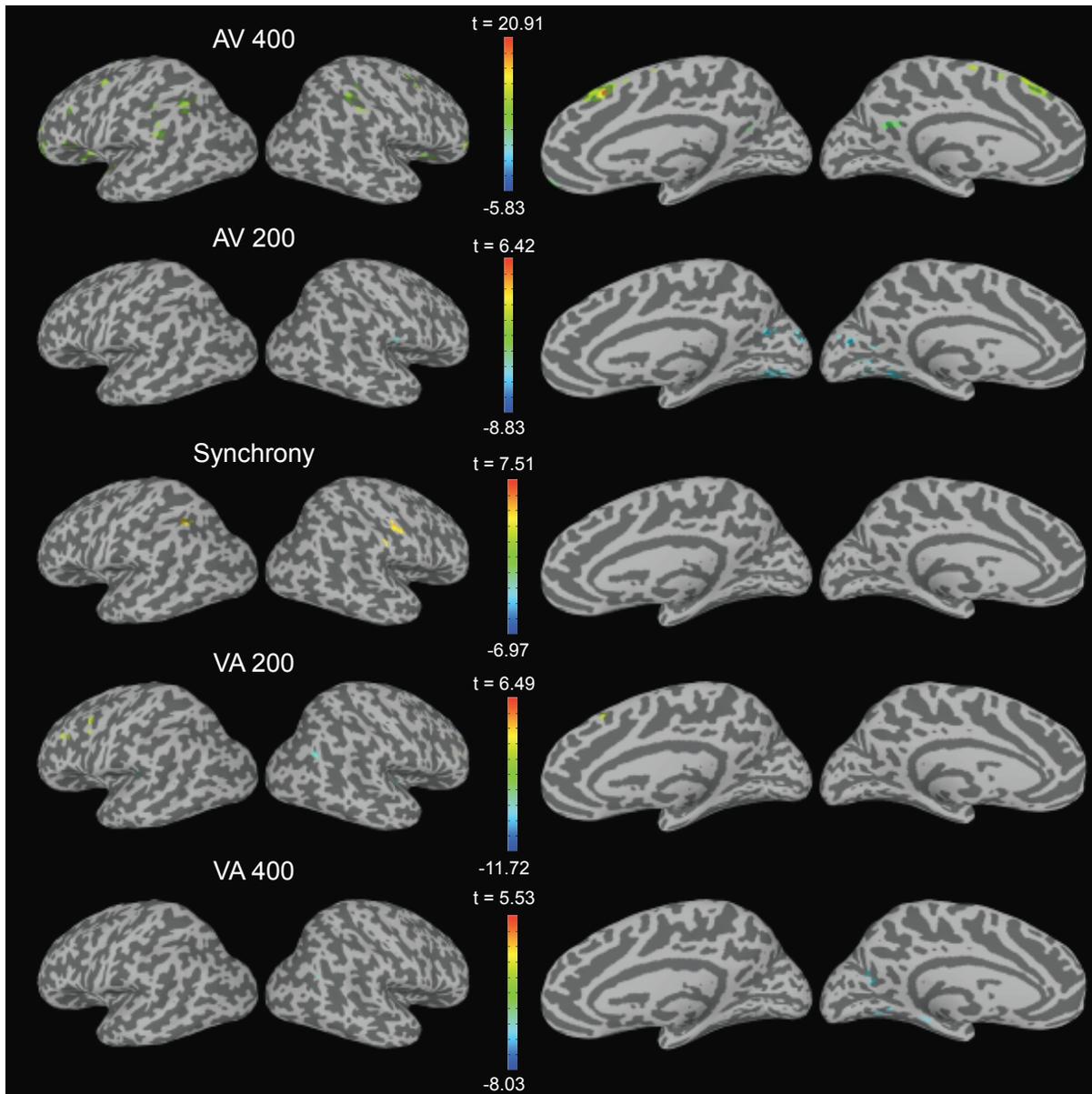


Figure 6-4. Simple main effect of training status at each SOA

				<i>Talairach Coordinates</i>		
				<i>Center of Mass</i>		
	<i>Voxels</i>	<i>Hemisphere</i>	<i>Structure</i>	<i>x</i>	<i>y</i>	<i>z</i>
Main Effect of Training Status	75	Right	Middle Temporal Gyrus	-45.6	73.1	26.8
	42	Right	Insula	-44.9	11.9	5.5
	41	Right	Medial Frontal Gyrus	-6	-18.1	44.8
	30	Right	Cingulate Gyrus	-20.5	-4.8	23.9
Main Effect of SOA	268	Right	Lingual Gyrus BA 18	-2.8	80.1	0
	164	Right	Supramarginal Gyrus BA 40	-55.9	44.8	25.9
	64	Left	Inferior Parietal Lobule BA 40	56.6	27.4	35.5
	57	Right	Lingual Gyrus	-24.4	89.6	-2.8
	55	Right	Middle Temporal Gyrus BA22	-62.2	47	4
	34	Left	Insula	24.2	-21.1	-7.1
	31	Right	Superior Frontal Gyrus BA 10	-40.3	-48.1	24.6
	29	Right	Superior Temporal Gyrus BA 38	-41.4	-16.2	-21.9

**Table 6-1. Effects of Training Status and SOA by Center of Mass** Whole brain analysis for main effects of Training Status (pre-training vs. post training) and SOA. Coordinates correspond to center of mass of cluster. Anatomical regions defined using Talairach atlas.

				<i>Talairach Coordinates</i>		
	<i>Voxels</i>	<i>Hemisphere</i>	<i>Structure</i>	<i>Peak</i>		
				<i>x</i>	<i>y</i>	<i>z</i>
Main Effect of Training Status	75	Right	Middle Temporal Gyrus	-49	73	22
	42	Right	Insula	-43	15	6
	41	Right	Medial Frontal Gyrus	-7	-19	44
	30	Right	Cingulate Gyrus	-23	-5	26
Main Effect of SOA	268	Right	Cuneus BA 18	-13	77	6
	164	Right	Supramarginal Gyrus	-53	47	28
	64	Left	Postcentral Gyrus BA 2	53	27	34
	57	Right	Lingual Gyrus	-25	87	-4
	55	Right	Middle Temporal Gyrus	-65	43	4
	34	Left	Subcallosal Gyrus BA 34	23	-21	-10
	31	Right	Middle Frontal Gyrus BA 9	-41	-47	26
	29	Right	Superior Temporal Gyrus BA 38	-41	-17	-22

**Table 6-2. Effects of Training Status and SOA by Peak Voxel** Whole brain analysis for main effects of Training Status (pre-training vs. post training) and SOA. Coordinates correspond to peak voxel of the cluster. Anatomical regions defined using Talairach atlas.

Talairach Coordinates

	Voxels	Hemisphere	Structure	<u>Center of Mass</u>		
				x	y	z
	583	Right	Superior Frontal Gyrus	-0.8	-9.9	48.5
	350	Left	Inferior Parietal Lobule	43.1	33.7	40.7
	231	Left	Red Nucleus	0.7	27.1	-10.2
	165	Right	Insula	-37.2	-17.6	1.9
	155	Left	Middle Frontal Gyrus	36.5	-31.8	21
	147	Right	Inferior Parietal Lobule	-41.2	41.5	44.5
	128	Left	Posterior Cingulate BA 31	1.8	54.7	22.5
	103	Left	Insula	28.5	-19.1	7.6
	91	Left	Culmen Dentate	15.7	50.6	-20.6
	89	Left	Precentral Gyrus BA 6	54.9	1.8	32.6
	84	Right	Middle Occipital Gyrus BA 18	-30.5	88.7	4.8
	83	Right	Precentral Gyrus Right BA 6	-30.1	8.3	50.6
	77	Left	Culmen	6	32.3	-23.1
	76	Right	Thalamus Anterior Nucleus	-9.1	12.2	14.3
	48	Left	Inferior Parietal Lobule	36.6	49.3	46.9
Training Status x SOA Interaction	47	Left	Left Insula BA13	46.2	38.2	17.7
	47	Left	Inferior Frontal Gyrus	39.4	-13	21.3
	47	Left	Postcentral Gyrus	54.3	19.8	27.2
	45	Left	Precentral Gyrus	47.8	-6.3	12.9
	44	Left	Inferior Occipital Gyrus	29.1	90.9	-6.2
	42	Right	Middle Frontal Gyrus	-40.2	-41.5	9
	42	Left	Middle Temporal Gyrus	46.6	75.2	23.6
	42	Left	Superior Frontal Gyrus	30.4	-39.9	30.6
	40	Left	Lingual Gyrus BA 17	12.7	94	-11.6
	40	Right	Inferior Frontal Gyrus BA 45	-53.5	-10.7	17.7
	38	Right	Postcentral Gyrus BA 40	-59	27	20.2
	36	Left	Middle Frontal Gyrus	34.5	-45.1	10.7
	33	Left	Precuneus BA 7	20.1	68.5	44.5
	32	Right	Middle Frontal Gyrus	-49.3	0.3	29.7
	30		Unknown	-7.2	30.9	-29.5
	29	Right	Paracentral Lobule	-1.2	20.3	47
	26	Left	Cingulate Gyrus	16.3	32.9	24.4
	26	Left	Precuneus	6.5	72.8	38.4
	25	Left	Thalamus Anterior Nucleus	9.2	12.6	13.6

**Table 6-3. Interaction of Training Status and SOA by Center of Mass** Whole brain analysis for the interaction of Training Status (pre-training vs. post training) and SOA. Coordinates correspond to center of mass of cluster. Anatomical regions defined using Talairach atlas.

*Talairach Coordinates*

	Voxels	Hemisphere	Structure	Peak		
				x	y	z
	583	Right	Superior Frontal Gyrus	-1	-5	54
	350	Left	Postcentral Gyrus	39	27	32
	231		<i>Superior Colliculus</i>	3	29	-12
	165	Right	Insula BA 13	-33	-21	4
	155	Left	Middle Frontal Gyrus	39	-31	24
	147	Right	Inferior Parietal Lobule	-47	33	42
	128	Right	Posterior Cingulate	-1	51	24
	103	Left	Insula	31	-9	10
	91	Left	Culmen	11	51	-18
	89	Left	Precentral Gyrus	53	3	30
	84	Right	Middle Occipital Gyrus	-27	89	6
	83	Right	Precentral Gyrus	-27	11	48
	77	Left	Culmen	1	31	-24
	76	Right	Thalamus	-9	17	10
			Medial Dorsal Nucleus			
	48	Left	Inferior Parietal Lobule BA 40	37	51	48
	47	Left	Superior Temporal Gyrus	49	37	16
	47	Left	Insula	37	-11	20
	47	Left	Postcentral Gyrus	55	19	26
Training Status	45	Left	Precentral Gyrus	47	-7	12
x	44	Left	Inferior Occipital Gyrus	31	91	-8
SOA	42	Right	Middle Frontal Gyrus	-41	-43	10
Interaction			BA 46			
	42	Left	Superior Occipital Gyrus BA 19	41	79	26
	42	Left	Superior Frontal Gyrus	31	-37	32
	40	Left	Lingual Gyrus BA 17	17	95	-10
	40	Right	Inferior Frontal Gyrus BA 44	-53	-11	18
	38	Right	Inferior Parietal Lobule BA 40	-59	27	22
	36	Left	Superior Frontal Gyrus	31	-45	14
	33	Left	Precuneus	21	67	42
	32	Right	Precentral Gyrus BA 6	-51	-1	30
	30		Unknown	-7	31	-30
	29	Right	Medial Frontal Gyrus BA 6	-3	17	48
	26	Left	Cingulate Gyrus	17	33	24
	26	Left	Precuneus	5	75	42
	25	Left	Thalamus	9	11	16
			Anterior Nucleus			

**Table 6-4. Interaction of Training Status and SOA by Peak Voxel** Whole brain analysis for the interaction of Training Status (pre-training vs. post training) and SOA. Coordinates correspond to peak voxel of the cluster. Anatomical regions defined using Talairach atlas. Superior colliculus labeled by author estimation.

Talairach Coordinates

	Voxels	Hemisphere	Structure	Center of Mass		
				x	y	z
	438	Right	Medial Frontal Gyrus	-2.2	-23.3	46
	125	Left	Inferior Frontal Gyrus	39	-36.2	9
	111	Left	Inferior Frontal Gyrus	44.3	-15.1	15.9
	79	Left	Angular Gyrus	32.2	53.8	38.4
	78	Left	Insula	30	-19.7	-0.3
	78	Right	Inferior Parietal Lobule	-31.8	43.7	37.5
	77	Left	Posterior Cingulate	2.5	54	22.5
	67	Right	Medial Frontal Gyrus	-0.9	-54.3	-3.8
	58	Right	Insula	-34.6	-20.7	1.1
	48	Left	Middle Frontal Gyrus	40	-11.4	34.5
			BA 9			
AV 400	45	Right	Inferior Parietal Lobule	-53.8	41.8	42.8
	44	Left	Supramarginal Gyrus	53.2	46.4	30.1
	41	Right	Middle Frontal Gyrus	-43.6	-33.4	15.3
	37	Right	Precentral Gyrus	-28.5	8.2	50.1
			BA 6			
	34	Right	Superior Frontal Gyrus	-0.7	-6.1	53.5
	34	Left	Medial Frontal Gyrus	2.3	9.9	57.6
	33	Left	Inferior Parietal Lobule	47.6	40.6	41.7
	32	Right	Culmen	-20.7	51.7	-19
	30	Left	Precentral Gyrus	27.7	9.2	53.2
	27	Left	Thalamus	16.7	23.9	10.1
			Pulvinar			
	25	Left	Insula	43.9	8.1	-3.7
	195	Right	Lingual Gyrus	-15	67.9	-6.1
			BA 18			
	87	Left	Cuneus	7.7	73.5	17.7
			BA 18			
	67	Left	Parahippocampal Gyrus	27.6	43.9	-5.5
			BA 19			
	56	Right	Precuneus	-22.1	62.5	19.2
	46	Left	Lingual Gyrus	18.5	75	-1.2
AV 200	42	Left	Culmen	8.4	43	-5.1
	31	Left	Lingual Gyrus	13.5	67.4	-5.1
	31	Right	Cuneus	-5.1	82.5	19.9
			BA 18			
	31	Right	Precentral Gyrus	-35.8	17.2	34.5
	27	Right	Inferior Parietal Lobule	-60.2	28.7	22.8
	26	Left	Cuneus	9.8	62.1	6.1
			BA 18			
	25	Right	Lingual Gyrus	-16	76.7	0.1
	84	Right	Postcentral Gyrus	-53.8	23.9	43.7
			BA 2			
Synchrony	41	Right	Middle Frontal Gyrus	-53.7	-2.7	42.2
	36	Left	Fusiform Gyrus	39.1	54.7	-15.9
	35	Left	Precuneus	26.5	51.6	38.2
	35	Left	Postcentral Gyrus	49.6	30.7	56

	28	Right	Superior Frontal Gyrus	-27.9	-38.7	9
	68	Right	Middle Temporal Gyrus BA 39	-44.3	73.2	28.5
	41	Right	Postcentral Gyrus	-29.6	19.8	28
	40	Left	Fusiform Gyrus BA 37	41.6	41	-13.5
	36	Right	Insula	-40.5	17.3	7.9
	36	Left	Middle Frontal Gyrus	50.1	-21.9	33.2
VA 200	28	Left	Superior Temporal Gyrus BA 41	42.1	34.4	16
	25	Left	Insula	33.1	-3.1	12.2
	25	Right	Middle Frontal Gyrus BA 9	-44.7	-31.4	30.3
	25	Left	Middle Frontal Gyrus BA9	40.9	-12.4	33.8
	25	Left	Precentral Gyrus	40.9	2	42.3
	25	Right	Medial Frontal Gyrus	-5.4	-19.1	47.3
	70	Left	Posterior Cingulate	18.6	58.2	16.4
	45	Left	Culmen	15.8	53.8	-7.7
VA 400	30	Left	Parahippocampal Gyrus	14.3	24.1	-11.4
	26	Right	Middle Temporal Gyrus	-47.9	72.2	23.1

**Table 6-5. Simple Main Effects Training Status at each SOA by Center of Mass** Whole brain analysis for simple main effects of Training Status (pre-training vs. post training) at each SOA. Coordinates correspond to center of mass of cluster. Anatomical regions defined using Talairach atlas.

Talairach Coordinates

	Voxels	Hemisphere	Structure	Peak		
				x	y	z
	438	Right	SMA	-5	-19	44
	125	Left	Inferior Frontal Gyrus	39	-35	2
	111	Left	Inferior Frontal Gyrus	51	-13	18
	79	Left	Angular Gyrus	31	53	38
	78	Left	Insula	31	-17	0
	78	Right	Cingulate Gyrus	-25	43	38
	77	Left	Cingulate Gyrus	7	55	26
			BA 31			
	67	Right	Medial Frontal Gyrus	-1	-61	-2
	58	Right	Inferior Frontal Gyrus	-33	-25	2
			BA 47			
	48	Left	Precentral Gyrus	37	-11	34
			BA 9			
	45	Right	Inferior Parietal Lobule	-57	39	42
			BA 40			
AV 400	44	Left	Supramarginal Gyrus	61	47	32
			BA 40			
	41	Right	Middle Frontal Gyrus	-43	-31	14
	37	Right	Precentral Gyrus	-31	9	52
	34	Right	Superior Frontal Gyrus	-1	-7	54
	34	Left	Medial Frontal Gyrus	3	11	58
	33	Left	Inferior Parietal Lobule	47	41	42
	32	Right	Culmen	-17	51	-18
			Dentate			
	30	Left	Precentral Gyrus	27	9	52
			BA 6			
	27	Left	Thalamus	17	23	10
			Pulvinar			
	25	Left	Insula	43	11	-2
			BA 13			
	195	Right	Lingual Gyrus	-15	69	-6
			18			
	87	Left	Cuneus	5	73	20
			BA 18			
	67	Left	Parahippocampal Gyrus	21	47	-6
	56	Right	Posterior Cingulate	-25	59	20
	46	Left	Lingual Gyrus	19	75	0
AV 200	42	Left	Culmen	9	43	-6
	31	Left	Lingual Gyrus	17	67	-6
			BA 19			
	31	Right	Cuneus	-3	83	20
	31	Right	Precentral Gyrus	-35	15	34
	27	Right	Inferior Parietal Lobule	-65	27	24
	26	Left	Cuneus	9	63	8
	25	Right	Lingual Gyrus	-17	79	-2
	84	Right	Postcentral Gyrus	-53	25	46
			BA 2			
Synchrony	41	Right	Precentral Gyrus	-53	-3	44
	36	Left	Fusiform Gyrus	43	51	-18

	35	Left	Supramarginal Gyrus	29	49	38
	35	Left	Postcentral Gyrus	47	29	56
			BA2			
	28	Right	Middle Frontal Gyrus	-29	-37	10
	68	Right	Middle Temporal Gyrus	-47	73	28
	41	Right	Insula	-29	19	28
	40	Left	Fusiform Gyrus	43	41	-16
			BA 37			
	36	Right	Insula	-41	17	8
	36	Left	Middle Frontal Gyrus	49	-25	34
VA 200	28	Left	Superior Temporal Gyrus	43	33	16
			BA 41			
	25	Left	Clastrum	31	-5	8
	25	Right	Middle Frontal Gyrus	-47	-29	32
	25	Left	Middle Frontal Gyrus	41	-11	34
			BA 9			
	25	Left	Precentral Gyrus	39	3	42
	25	Right	SMA	-3	-19	46
	70	Left	Posterior Cingulate	23	55	18
	45	Left	Culmen	17	55	-8
VA 400	30	Left	Parahippocampal Gyrus	19	23	-12
			BA 35			
	26	Right	Middle Temporal Gyrus	-47	73	24

**Table 6-6. Simple Main Effects Training Status at each SOA by Peak Voxel** Whole brain analysis for simple main effects of Training Status (pre-training vs. post training) at each SOA. Coordinates correspond to peak voxel of the cluster. Anatomical regions defined using Talairach atlas.

				<i>Talairach Coordinates</i>		
				<u>Center of Mass</u>		
	<i>Voxels</i>	<i>Hemisphere</i>	<i>Structure</i>	<i>x</i>	<i>y</i>	<i>z</i>
Main Effect of Training Status	75	Right	Angular Gyrus	-45.6	73.1	26.8
		Right	Anterior Transverse Temporal Gyrus (of Heschl)	-44.9	11.9	5.5
	42	Right	Superior Frontal Gyrus	-6	-18.1	44.8
	41	Right	Cerebral White Matter	-20.5	-4.8	23.9
	30	Right	Cerebral White Matter	-20.5	-4.8	23.9
Main Effect of SOA	268	Right	Cerebral White Matter	-2.8	80.1	0
			<i>Cuneus</i>			
			<i>Lingual Gyrus</i>			
		Left	<i>Cuneus</i>			
	164	Right	Cerebral White Matter	-55.9	44.8	25.9
			<i>Angular Gyrus</i>			
			<i>Supramarginal Gyrus</i>			
	64	Left	Supramarginal Gyrus	56.6	27.4	35.5
	57	Right	Cerebral White Matter	-24.4	89.6	-2.8
			<i>Occipital Pole</i>			
55	Right	Middle Temporal Gyrus	-62.2	47	4	
34	Left	Cerebral White Matter	24.2	-21.1	-7.1	
31	Right	Frontal Middle Gyrus	-40.3	-48.1	24.6	
29	Right	Temporal Pole	-41.4	-16.2	-21.9	

**Table 6-7. Effects Training Status and SOA by Center of Mass and Defined by Author** Whole brain analysis for main effects of Training Status (pre-training vs. post training) and SOA. For larger clusters, nearest areas shown in italics as estimated by author. Non-italicized labels correspond to labels from atlas from (Destrieux, Fischl, Dale, & Halgren, 2010).

	Voxels	Hemisphere	Structure	Center of Mass		
				x	y	z
	583	Left	Superior Frontal Gyrus	-0.8	-9.9	48.5
		Right	<i>Superior Frontal Gyrus</i>			
	350	Left	Postcentral Sulcus	43.1	33.7	40.7
	231		<i>Superior Colliculus</i>	0.7	27.1	-10.2
	165	Right	Short Insular Gyri	-37.2	-17.6	1.9
	155	Left	Cerebral White Matter	36.5	-31.8	21
			<i>Inferior Frontal Sulcus</i>			
	147	Right	Cerebral White Matter	-41.2	41.5	44.5
	128	Left	Precuneus	1.8	54.7	22.5
	103	Left	Cerebral White Matter	28.5	-19.1	7.6
			<i>Circular Sulcus of the Insula</i>			
			<i>Short Insular Gyri</i>			
	91	Left	Cerebellum	15.7	50.6	-20.6
	89	Left	Precentral Gyrus	54.9	1.8	32.6
	84	Right	Middle Occipital Gyrus	-30.5	88.7	4.8
	83	Right	Cerebral White Matter	-30.1	8.3	50.6
			<i>Precentral Gyrus</i>			
	77		Unknown	6	32.3	-23.1
	76	Right	Thalamus	-9.1	12.2	14.3
	48	Left	Cerebral White Matter	36.6	49.3	46.9
			<i>Superior Parietal Lobule</i>			
Training Status	47	Left	Cerebral White Matter	46.2	38.2	17.7
x	47	Left	Cerebral White Matter	39.4	-13	21.3
SOA	47	Left	Cerebral White Matter	54.3	19.8	27.2
Interaction	45	Left	Cerebral White Matter	47.8	-6.3	12.9
			<i>Opercular part of the Inferior Frontal Gyrus</i>			
	44	Left	Occipital Pole	29.1	90.9	-6.2
	42	Right	Cerebral White Matter	-40.2	-41.5	9
			<i>Middle Frontal Gyrus</i>			
	42	Left	Angular Gyrus	46.6	75.2	23.6
	42	Left	Middle Frontal Gyrus	30.4	-39.9	30.6
	40	Left	Occipital Pole	12.7	94	-11.6
	40	Right	Opercular part of the Inferior Frontal Gyrus	-53.5	-10.7	17.7
	38	Right	Supramarginal Gyrus	-59	27	20.2
	36	Left	Cerebral White Matter	34.5	-45.1	10.7
			<i>Middle Frontal Gyrus</i>			
	33	Left	Superior Parietal Lobule	20.1	68.5	44.5
	32	Right	Cerebral White Matter	-49.3	0.3	29.7
	30		Unknown	-7.2	30.9	-29.5
	29	Right	Superior Frontal Gyrus	-1.2	20.3	47
	26	Left	Cerebral White Matter	16.3	32.9	24.4
	26	Left	Cerebral White Matter	6.5	72.8	38.4
			<i>Precuneus</i>			
			<i>Superior Parietal Lobule</i>			
	25	Left	Thalamus	9.2	12.6	13.6

**Table 6-8. Interaction of Training Status and SOA by Center of Mass and Defined by Author**  
 Whole brain analysis for interaction of Training Status (pre-training vs. post training) and SOA. For larger clusters, nearest areas shown in italics as estimated by author. Superior colliculus labeled by author estimation. Non-italicized labels correspond to labels from atlas from (Destrieux et al., 2010).

Talairach Coordinates

<i>Voxels</i>	<i>Hemisphere</i>	<i>Structure</i>	<i>Center of Mass</i>			
			<i>x</i>	<i>y</i>	<i>z</i>	
AV 400	438	Right	Superior Frontal Gyrus	-2.2	-23.3	46
		Left	Superior Frontal Gyrus			
	125	Left	Cerebral White Matter	39	-36.2	9
	111	Left	Cerebral White Matter	44.3	-15.1	15.9
			<i>Opercular part of the Inferior Frontal Gyrus</i>			
	79	Left	Intraparietal Sulcus and Transverse Parietal Sulci	32.2	53.8	38.4
	78	Left	Circular Sulcus of the Insula	30	-19.7	-0.3
	78	Right	Cerebral White Matter	-31.8	43.7	37.5
	77	Left	Precuneus	2.5	54	22.5
	67	Left	Gyrus Rectus	-0.9	-54.3	-3.8
	58	Right	Circular Sulcus of the Insula	-34.6	-20.7	1.1
	48	Left	Cerebral White Matter	40	-11.4	34.5
	45	Right	Supramarginal Gyrus	-53.8	41.8	42.8
	44	Left	Angular Gyrus	53.2	46.4	30.1
	41	Right	Cerebral White Matter	-43.6	-33.4	15.3
	37	Right	Cerebral White Matter	-28.5	8.2	50.1
	34	Right	Superior Frontal Gyrus	-0.7	-6.1	53.5
	34	Left	Superior Frontal Gyrus	2.3	9.9	57.6
	33	Left	Supramarginal Gyrus	47.6	40.6	41.7
	32	Right	Cerebellum	-20.7	51.7	-19
30	Left	Cerebral White Matter	27.7	9.2	53.2	
27	Left	Thalamus	16.7	23.9	10.1	
25	Left	Cerebral White Matter	43.9	8.1	-3.7	
AV 200	195	Right	Cerebral White Matter	-15	67.9	-6.1
			<i>Lingual Gyrus</i>			
	87	Left	Cerebral White Matter	7.7	73.5	17.7
			<i>Cuneus</i>			
	67	Left	Collateral Sulcus and Lingual Sulcus	27.6	43.9	-5.5
	56	Right	Cerebral White Matter	-22.1	62.5	19.2
	46	Left	Cerebral White Matter	18.5	75	-1.2
	42	Left	Cerebellum	8.4	43	-5.1
	31	Left	Cerebral White Matter	13.5	67.4	-5.1
	31	Right	Cuneus	-5.1	82.5	19.9
	31	Right	Cerebral White Matter	-35.8	17.2	34.5
27	Right	Supramarginal Gyrus	-60.2	28.7	22.8	
26	Left	Cerebral White Matter	9.8	62.1	6.1	
25	Right	Cerebral White Matter	-16	76.7	0.1	
Synchrony	84	Right	Postcentral Gyrus	-53.8	23.9	43.7
	41	Right	Precentral Gyrus	-53.7	-2.7	42.2
	36	Left	Lateral Occipito-temporal Gyrus	39.1	54.7	-15.9
	35	Left	Cerebral White Matter	26.5	51.6	38.2
	35	Left	Unknown	49.6	30.7	56
28	Right	Cerebral White Matter	-27.9	-38.7	9	
VA 200	68	Right	Angular Gyrus	-44.3	73.2	28.5
	41	Right	Cerebral White Matter	-29.6	19.8	28
	40	Left	Cerebral White Matter	41.6	41	-13.5

	36	Right	Anterior Transverse Temporal Gyrus (of Heschl)	-40.5	17.3	7.9
	36	Left	Middle Frontal Gyrus	50.1	-21.9	33.2
	28	Left	Posterior ramus of the Lateral Sulcus	42.1	34.4	16
	25	Left	Circular Sulcus of the Insula	33.1	-3.1	12.2
	25	Right	Middle Frontal Gyrus	-44.7	-31.4	30.3
	25	Left	Cerebral White Matter	40.9	-12.4	33.8
	25	Left	Cerebral White Matter	40.9	2	42.3
	25	Right	Superior Frontal Gyrus	-5.4	-19.1	47.3
	70	Left	Cerebral White Matter	18.6	58.2	16.4
			<i>Parieto-occipital Sulcus</i>			
VA 400	45	Left	Parieto-occipital Sulcus	15.8	53.8	-7.7
	30	Left	Unknown	14.3	24.1	-11.4
	26	Right	Angular Gyrus	-47.9	72.2	23.1

**Table 6-9. Simple Main Effects Training Status at each SOA by Center of Mass and Defined by Author** Whole brain analysis for simple main effects of Training Status (pre-training vs. post training) at each SOA. For larger clusters, nearest areas shown in italics as estimated by author. Non-italicized labels correspond to labels from atlas from (Destrieux et al., 2010).

## References

- Argall, B. D., Saad, Z. S., & Beauchamp, M. S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. *Human Brain Mapping, 27*(1), 14-27.
- Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339-355.
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14*(2), 427-438.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*, 162-173.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179-194.
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage, 53*(1), 1-15.
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics, 66*(8), 1388-1404.
- Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision research, 39*(19), 3197-3221.
- Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America, 95*(23), 13988-13993.  
doi:10.1073/pnas.95.23.13988
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9*(2), 195-207.

- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *33*(5), 636-647.
- Kahnt, T., Grueschow, M., Speck, O., & Haynes, J.-D. (2011). Perceptual Learning and Decision-Making in Human Medial Frontal Cortex. *Neuron*, *70*(3), 549-559.  
doi:10.1016/j.neuron.2011.02.054
- Marchant, J. L., Ruff, C. C., & Driver, J. (2012). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target - detection performance for both modalities. *Human brain mapping*, *33*(5), 1212-1224.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., . . . Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356*(1412), 1293-1322.
- Petrov, A. A., Doshier, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological review*, *112*(4), 715.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, *32*(18), 6263-6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, *29*(39), 12265-12274.
- Powers III, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports*, *6*, 23374.

- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147-1153.
- Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), B69-B78.
- Skudlarski, P., Constable, R. T., & Gore, J. C. (1999). ROC analysis of statistical methods used in functional MRI: individual subjects. *NeuroImage*, *9*(3), 311-329.  
doi:10.1006/nimg.1999.0402
- Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, *49*(4), 3308-3318.
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*, *34*(3), 691-697.
- Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., & James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*, *55*(3), 1339-1345.
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. *Journal of Experimental Psychology-Human Perception and Performance*, *38*(6), 1517-1529. doi:10.1037/a0027339
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26*(2), 212-215.
- Van der Burg, E., Cass, J., Olivers, C. N., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *Plos One*, *5*(5), e10664.

- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & psychophysics*, 69(5), 744-756.
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64, 105-123.
- Williams, J. T., Darcy, I., & Newman, S. D. (2016). Bimodal bilingualism as multisensory training?: Evidence for improved audiovisual speech perception after sign language exposure. *Brain Research*, 1633, 101-110.
- Xiong, J., Gao, J., Lancaster, J., & Fox, P. (1995). Clustered pixels analysis for functional MRI activation studies of the human brain. *Human Brain Mapping*, 3(4), 287-301.  
doi:10.1002/hbm.460030404

## Chapter VII

### Discussion

In this section, the results and implications of the previous chapters will be summarized, pertinent themes related to multisensory perceptual learning will be examined, and future investigations, directions, and considerations regarding multisensory perceptual learning will be discussed.

#### Summary of Findings

##### *Multisensory Perceptual Learning is Dependent on Feedback*

The findings of Chapters 2 and 5 suggest that an objective feedback signal is necessary for multisensory perceptual learning to occur. Experiment 1 in Chapter 2 finds that temporal acuity is only significantly enhanced when an objective feedback signal is present during perceptual training. When false feedback is presented, even on only 20% of trials, during perceptual training, an enhancement of multisensory temporal acuity is not observed. Furthermore, increasing experience with passive exposure to either “flash-beep” stimuli (Chapter 2), similar to Powers *et al.* (2009), or audiovisual speech stimuli (Chapter 5) does not enhance temporal acuity and may even widen the TBW (Powers, Hillock, & Wallace, 2009). Additionally, practice (i.e. completing the SJ task) in the absence of feedback does enhance temporal acuity for audiovisual speech. Thus, these findings collectively suggest that perceptual training to enhance multisensory temporal acuity requires a feedback signal to be present during perceptual training.

Chapter 3 further suggests that an objective feedback signal during perceptual training is not always sufficient to elicit perceptual learning but rather the stimuli presented during perceptual training must be sufficiently difficult. This implies that feedback must be informative

such that an individual is presented with negative feedback rather than feedback that confirms perceptual responses for relatively easy stimuli. This suggests that some of the variability observed for the training protocols in Chapters 2, 4, 5, and 6 might be driven by relative differences in difficulty of the training task for each individual. Thus, the correlation between the size of the TBW prior to perceptual training and the magnitude of the change in the TBW following perceptual training (as observed in Chapter 2) may also be a result of relative differences in difficulty of the perceptual training task.

Finally, Experiment 2 in Chapter 2 suggests that although feedback signals may be necessary for perceptual training to enhance temporal acuity, numerous types of feedback signals (symbolic, monetary, and social feedback) are each capable of driving substantial changes in temporal acuity following perceptual training.

#### *Multisensory Perceptual Training Enhances Temporal Acuity for Audiovisual Speech*

Chapter 5 demonstrates that multisensory perceptual training is capable of enhancing temporal acuity for more complex multisensory stimuli like audiovisual speech. Previously, studies of multisensory perceptual learning had only reported that perceptual training enhanced multisensory temporal acuity for “flash-beep” like stimuli. Although perceptual training enhances temporal acuity for audiovisual speech, the enhancement of temporal acuity appears to be less durable for audiovisual speech compared to “flash-beep” stimuli.

The findings of Chapter 5 also suggest, however, that perceptual training does not generalize across levels of stimulus complexity such that perceptual training with “flash-beep” or audiovisual speech stimuli did not enhance temporal acuity for audiovisual speech or “flash-beep” stimuli respectively. Although some prior studies have observed generalized improvement in temporal acuity for unisensory presentations of the same type of stimulus (Powers III, Hillock-Dunn, & Wallace, 2016) or a change in the perception of a multisensory illusion (e.g. decreased perception of the SIFI), generalization across different types of stimuli has not been reported.

Thus, it is possible that multisensory perceptual learning elicits changes in temporal perception that are specific to certain stimuli.

*Perceptual Training with Audiovisual Speech Stimuli Elicits Changes in a Wider Network of Multisensory and Decision-making Areas*

Chapter 6 provides preliminary evidence that perceptual training with audiovisual speech stimuli elicits changes in a wider network of higher order cortical regions in comparison to previous reports of multisensory perceptual learning that had employed “flash-beep” stimuli. These findings suggest that perceptual training may enhance temporal acuity through changes in activity at higher-order structures as well as sensory cortices rather than the sensory representation. Additionally, changes in neural activity following perceptual training may vary greatly across different SOAs.

*Feedback Influences the Time Course of Cumulative and Rapid Multisensory Temporal Recalibration*

In Chapter 4, more robust temporal recalibration is observed for individuals completing an SJ task with feedback in comparison to completing the SJ task in the absence of feedback. We also observe different time courses for cumulative and rapid temporal recalibration dependent upon the type of feedback provided on the previous trial such that cumulative recalibration relied on prior feedback that informed perceptual judgments (i.e. negative feedback) while rapid recalibration relied more on feedback that confirmed perceptual judgments (i.e. positive feedback). Furthermore, the time course of temporal recalibration was altered by the reliability of the feedback provided. When uncorrelated feedback (i.e. false feedback was provided on 50% of the trials) was removed and subjects continued to complete the task in the absence of feedback, greater rapid temporal recalibration was observed suggestive that these individuals were more likely to weight immediate sensory experience to

recalibrate their audiovisual temporal representation. Collectively, these findings are suggestive that cumulative and rapid recalibration may function via distinct mechanisms.

### **Relevant Themes Regarding Multisensory Temporal Perception and Plasticity**

This section will discuss relevant topics regarding how we conceptualize multisensory processing and the potentially unique aspects of multisensory perceptual learning.

#### *Interaction of Top-down and Bottom-up Processing*

Although the interaction between sensory and cognitive processes is a general theme across all Chapters, the findings of Chapter 4 in particular suggest that multisensory plasticity likely occurs at different level of the cortical hierarchy. Here, we report that feedback signals distinctly influence the time course of what are likely two separate forms of multisensory temporal plasticity (i.e. rapid and cumulative temporal recalibration). A similar study also observed separate (and possibly distinct) time courses of temporal recalibration such that following a period of adaptation to an asynchronous stimulus, there is a cumulative change in the PSS to return towards the baseline PSS but trial-to-trial recalibration of the PSS (i.e. rapid recalibration) remains constant following adaptation (Van der Burg, Alais, & Cass, 2015). The dissociation of the time courses of temporal recalibration following an adaptation period were suggested by *Van der Burg et al. (2015)* to possibly reflect different stages of the sensory-decisional process, with changes in temporal recalibration following adaptation resulting from reweighting of sensory evidence at a more cognitive, higher-level decisional stages whereas rapid temporal recalibration might operate a more sensory processing levels (Van der Burg et al., 2015). This is further supported by evidence that rapid temporal recalibration is also evident for audiovisual speech and not dependent whether the same speaker was presented on the previous trial or whether the gender of the visual and auditory signals were congruent on the prior trial (Van der Burg & Goodbourn, 2015).

The initial findings in Chapter 4 appear to support the conclusions of *Van der Burg et al. (2015)* as we report that in the absence of feedback, the time course of cumulative temporal recalibration results, at least for a period of trials, results in a change in the PSS while the time course of rapid temporal recalibration remains constant such that the trial-to-trial change in the PSS remains the same. Yet following the introduction of a feedback signal, the time course of rapid temporal recalibration also changes dependent upon the prior feedback (i.e. correct or incorrect feedback) as well as the reliability of the feedback signals. Thus, while cumulative and rapid temporal recalibration appear to exhibit distinct timecourse in the absence of feedback, when feedback signals are present the time course of rapid recalibration appears to also be influenced by presumably top-down feedback signals. These findings suggest that while rapid temporal recalibration may be driven by sensory processes, top-down signals may interact with sensory areas to elicit multisensory temporal recalibration.

Chapter 6 also provides preliminary evidence that multisensory perceptual learning may result from changes in both sensory representations and changes in decision-making areas. In Chapter 6, activity in numerous regions in frontal and parietal cortices is altered following perceptual training to enhance temporal acuity for audiovisual speech. Evidence from visual perceptual learning suggests regions of frontal and parietal cortices as well as sensory cortex are important for perceptual learning (Kahnt, Grueschow, Speck, & Haynes, 2011). Additionally, frontal cortical regions have also recently been suggested to be an important region for the interaction of top-down and bottom-up multisensory processes (Gau & Noppeney, 2016). Thus, it is possible that the changes in activity we observe in Chapter 6 reflect an interaction of top-down and bottom-up sensory processes.

Interestingly, the changes we observe in Chapter 6 differ from initial descriptions of multisensory perceptual learning that reported changes primarily in sensory areas as well as multisensory areas such as the pSTS (Powers, Hevey, & Wallace, 2012). It is possible that speech, unlike “flash-beep” stimuli, may exhibit structural specialization for the perception of

speech that is constructed in a way such that feedforward and feedback signals optimize speech perception. Learned multisensory associations appear to develop and optimize other perceptual processes, like facial recognition, that exhibit cortical specificity. For example, a study by *Von Kriegstein & Giraud (2006)* demonstrated that that learned multisensory associations of ecologically valid and redundant stimuli (e.g. voices associated with faces in contrast to ring tones associated with cell phones) facilitates unisensory object recognition due to the refinement of the predictive signals generated across sensory systems. In this study, the authors observed using fMRI enhanced functional connectivity between the fusiform face area (FFA) and temporal voices areas for faces paired with voices after multisensory learning that was also associated with enhanced speaker recognition facilitation for associated voice face pairs but not for associated voice name pairs. As cell phones associated with ring tones or brand names elicited less facilitation of object recognition compared to faces paired with voices and activity in the FFA was only enhanced by voices that had been paired with faces in contrast to voices paired with names, suggests that speaker recognition templates are enhanced by visual features that allow for better predictive representations of speaker identity (*Von Kriegstein & Giraud, 2006*). In the context of the findings in Chapter 6, the neural regions important to the perception of speech, an ethologically relevant stimulus, may enhance the capacity to perceive temporal synchrony.

### *Integrating the Principles of Multisensory Integration*

Though the temporal principle of multisensory integration pertains to many of the previously discussed findings in Chapter 2-6, it remains likely that other aspects of multisensory integration beyond temporal proximity affect multisensory perceptual learning and may underlie some of the improvements in temporal acuity. As described in the introduction, temporal perception is also influenced by spatial factors (*Spence, Baddeley, Zampini, James, & Shore, 2003; Zampini, Shore, & Spence, 2003*), relative stimulus effectiveness (*Fister, Stevenson,*

Nidiffer, Barnett, & Wallace, 2016), and semantic relationships and learned associations (Doehrmann & Naumer, 2008; Tsilionis & Vatakis, 2016) (in addition to a variety of attributes related to each unisensory stimulus [e.g. the brightness of the visual stimulus]). The similarity of the temporal principle and the other principles of multisensory integration is that each principle describes the relative probability that the separate sensory cues may be attributed to a common source or event and thus should be perceptually bound. Thus, it is not surprising that each of these principles interact to influence temporal acuity. It may be interesting and beneficial to consider that the principles of multisensory integration describe the extent to which these sensory (or amodal) stimulus characteristics predict whether separate sensory cues actually should be associated (i.e. go together)<sup>1</sup>.

---

<sup>1</sup> It is interesting to consider to what extent each principle of multisensory processing develops due to experience or if the principle exists *a priori*. As described earlier, evidence from the cat model suggests that spatial relationships between sensory stimuli do not exist *a priori* but develop through experience, particularly for temporally coincident stimuli (Wallace & Stein, 2007). Thus, the temporal relatedness of the stimuli appears to be a primary amodal feature that allows for the initial development of multisensory integration. At some level, temporally coincident sensory stimuli, at least for a moment, form a common perceptual experience and therefore the default nature of perception may be to ascribe temporally related features to a common source until other features suggest otherwise. A parallel consideration is what constitutes a moment of perceptual experience (i.e. if a TBW exists). A moment may constitute both a period at the neural and perceptual to account for either neural and physical conduction differences or the temporal limits of conscious perception respectively. Neural evidence from the cat model, seems to suggest that multisensory integration initially appears only for coincident stimuli, prior to the development of a TBW (Wallace & Stein, 1997). The temporal principle (and

Therefore, rather than considering separate multisensory principles, it may be more apt to instead consider the factors, both sensory and cognitive, that contribute to form a representation of whether stimuli should be associated or discriminated. In this context, past experience and learned associations as well as the statistical regularities of the sensory environment may each be equated in terms of how well they predict the association of sensory cues. When considering perceptual learning, temporal representations are likely updated and influenced by the predictive capacity of these stimulus attributes, which may be more readily apparent when feedback signals are available. Although it may appear that spatial and temporal coincidence are inherently predictive, as multisensory processing develops and matures, the predictive relationship of sensory signals may become more complex and thus all multisensory stimuli may not elicit multisensory facilitation. For example, neural evidence suggests that in some instances multisensory interactions are only evident for that visual content that predicts the timing of an auditory event (e.g. the clapping of hands) as similar multisensory interactions are not observed for audiovisual stimuli that lack predictive visual content (e.g. tearing a piece of paper) (Stekelenburg & Vroomen, 2007). It is likely then that predictive elements of sensory experience affect both sensory and cognitive levels of the neural hierarchy such that multisensory interactions may become more specific as well as task dependent. In the context of the current studies, temporal acuity may be enhanced by increased salience of particular stimulus features that are predictive of synchrony.

In the studies described in the previous chapters, the predictive nature of sensory cues is most likely to be evident for the audiovisual speech stimuli in Chapters 5 and 6. Speech is a particular class of stimulus with which we have extensive experience as well as informative prior knowledge. The visual component of speech, particularly the movements of the mouth, contains

---

TBW) may develop from experience yet temporal coincident may be the fundamental feature for the development of initial multisensory interactions.

numerous cues that are predictive of the temporal structure of the auditory speech component, including the temporal delay between the sensory cues (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009). Furthermore, semantic congruency is known to influence temporal acuity of audiovisual speech, such that asynchrony is more readily apparent for semantically incongruent speech (Vatakis & Spence, 2007). Each of these factors is predictive of whether these sensory signals should be perceptually bound. However, the predictive advantage conferred by these signals is not likely for the purpose of facilitating synchrony judgments but rather to enhance speech intelligibility through meaningful multisensory interactions. Indeed, the speech the visual component of speech (that precedes the auditory signal) is hypothesized to constrain the expected auditory signal based on phonological knowledge and subsequently influence the intelligibility of the speech signal through the interaction of these multisensory signals at a neural level (Virginie van Wassenhove, Grant, & Poeppel, 2005). Yet in the case of making a synchrony judgment, it is possible that following perceptual training particular aspects of the representation of the temporal relationship for the specific speech cue are amplified in order to better recognize the synchrony of the cues rather than predict the phonemic content of speech.

The dynamic motion cues of visual speech are known to enhance speech comprehension (Jaekl, Pesquita, Alsius, Munhall, & Soto-Faraco, 2015). Thus, in the case of the current studies, enhanced binding of particular motion cues for visual speech that are predictive of the auditory speech signal following perceptual training may result in a narrower TBW. In particular, visual cues that predict the auditory signal are most likely to enhance temporal acuity. One study that examined the effect of predictive and postdictive (i.e. after the auditory stimulus had occurred) information on temporal perception as measured by SJ and TOJ tasks observed that for the SJ task, the slope of the auditory-leading portion of the SJ curve was steeper when predictive information was available. Furthermore, the PSS estimated by TOJ task was shifted closer to the auditory-leading portion of the SJ curve when predictive

information was available such that the PSS from the TOJ was closer to true synchrony. As the stimulus for this task was a pendular motion that predicted the impact of the visual stimulus with another visual stimulus such that the synchronous sound would be expected at the point of impact, the results suggest that predictive information that suggests when the sound should occur leads to a steeper slope of the auditory-leading portion of the SJ curve and a PSS on the TOJ task that is close to true synchrony (Van Eijk, Kohlrausch, Juola, & Van De Par, 2010). Evidence also suggests that multisensory interactions may arise for artificial stimuli if the relationship between the auditory and visual stimuli is predictable. At the neural level, multisensory interactions are observed for the synchronous onset of a tone and the contact of two moving disks that result in the deformation of a rectangle. Yet no multisensory interactions are observed when the moving disks were not present, the collision of the disks and the rectangle was asynchronous, or there was no predictable relationship between the disks contacting the rectangle and the tone (i.e. asynchronous and synchronous trials varied on a trial-to-trial basis rather than being presented in a block of either synchronous or asynchronous stimuli)(Vroomen & Stekelenburg, 2010). Thus, in the context of the current studies, the motion cues present in visual speech that are most predictive of synchrony may be amplified by perceptual training and may result in greater multisensory interaction as salient cues that predict the auditory signal become more evident.

Some cues, in particular semantic relationships, that predict whether separate sensory cues should be associated might actually favor a wider TBW. As described in the introduction, the JND or TBW is often reported to be greater for semantically congruent speech compared to semantically incongruent speech (Virginie van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2007). Thus, some predictive cues may also promote a greater tolerance for asynchrony. Alternatively, as some suggest, the perception of synchrony and temporal binding may be related but dissociable and thus perceptual binding may still occur despite the perception of asynchrony (Tsilionis & Vatakis, 2016; Virginie van Wassenhove et al., 2007). For

speech the statistical regularities in the delay of the auditory signal relative to visual cues may be a learned association that promotes tolerance for asynchrony. In the context of the current studies, it is possible that the expectation of audiovisual asynchrony for speech cues supersedes perceptual learning that enhances temporal acuity and thus leads to an increase in the TBW in the week following perceptual training. Interestingly, even for more simple stimuli, the TBW may be increased when these associations between the auditory and sensory cues are predictable or are expected to be congruent, although the increase in tolerance might be modulated by expectations generated by the individual rather than the sensory environment. A study recently demonstrated that learned sensory outcomes that result from individual actions are more likely to be perceived as synchrony over a wider range of SOAs (i.e. the TBW for these stimuli was larger) compared to novel stimulus relationships that were not predicted by the individual's action. In comparison, stimulus relationships that were predicted by sensory cues were perceived to be synchronous over the same range of SOAs as those stimuli that were not predicted by sensory cues (Desantis & Haggard, 2016). Although the temporal relationships were not perceived to be generated by the participants in the present studies, it is interesting to consider that perceptual binding may actually default to decreasing temporal acuity for audiovisual pairs with increasing experience.

The most basic stimulus used by the studies described in the previous chapters is the “flash-beep” stimulus, a stimulus that is static, devoid of most spatial properties, and likely lacking any semantic relationship suggesting that this particular “flash” and tone should be associated. Yet for this simple stimulus, we observe that prior sensory experience influences the temporal representation in direction consistent with the temporal relationships of prior stimuli (i.e. temporal recalibration, see Chapter 4). Thus, if the prior stimulus had an auditory-lead, the PSS is adjusted in the direction such that the new representation of this stimulus will better account for an auditory-lead as experience predicts an auditory leading stimulus. Such might reflect a sensory driven adjustment in the perceptual representation, although it is also possible

that temporal recalibration is actually a higher-order perceptual adjustments. A recent study demonstrated the effects of prior stimulus history and test stimulus predictability for an adaptation experiment using “flash-beep” like stimuli. When participants were not aware of what test stimuli they would be asked to judge the simultaneity of (i.e. asked to make a retrospective SJ judgment), the study observed greater perception of asynchrony following adaptation with prior stimulus sequences that were either progressively synchronizing or at a constant SOA of 150 ms (audio- and visual-leading) in comparison to prior stimulus sequences that were desynchronizing or presented at synchrony or in the absence of adaptation sequences. In contrast, adaptation effects were not observed when the test stimuli were predictable (i.e. prospective SJ judgments). These effects suggest that the adaptation effects are the result of effects a decisional levels as differences were observed on the basis of retrospective versus prospective judgments. It was hypothesized that when test stimulus is unpredictable and at the end of an adaptation sequence, individuals may continuously updated their prior for audiovisual synchrony, which influenced the participants’ judgment of the simultaneity of the test stimulus, while when the test stimulus was predictable, participants may have made an isolated judgment of simultaneity independent of prior experience (Martin, Kösem, & Van Wassenhove, 2015). Similar to this hypothesis, it is possible that the SJ made by subjects, particularly in Chapter 4, would rely on basic representation of audiovisual relationships when judging the simultaneity of the relatively novel and basic “flash-beep” stimulus but are affected by cumulative sensory experience after repeated synchrony judgments.

#### *Generalizability of Multisensory Perceptual Training*

An important finding of Chapter 5 was that changes in temporal acuity did not generalize across different levels of stimulus complexity. One possible explanation is that the lack of generalization observed for the experiments in Chapter 5 are related to the specificity of perceptual representations in either the visual or auditory systems, which might limit the

generalization of a multisensory feature. Interestingly, perceptual learning may enhance temporal acuity for auditory stimuli but the increase in temporal acuity has been observed to be both specific and likely not exhibit generalization. This is supported by evidence from a study by *Mossbridge et al. (2006)* that sought to examine temporal perception following auditory perceptual learning. In this study, participants were either trained (with feedback) on a 2-IFC task to detect whether onset of two tones of different frequencies occurred asynchronously (i.e. whether the tones began at the same time or different times) or detect the order of the onset of two tones (i.e. in one interval one tone frequency led while in the other interval the other tone frequency led) or did not complete any training. Training resulted in task specific improvements, either for asynchrony training or order training, that was greater than the improvements observed for controls (controls performance improved following the post-test, however, the improvements in performance for the trained participants was greater than the improvement for controls). No generalization of improvements following training was observed to generalize across tasks or for the trained task with tones of different frequencies than the tones presented during training. These findings suggests that changes following auditory perceptual learning are very specific (and that asynchrony detection and temporal order detection may be separate processes, see Introduction) (*Mossbridge, Fitzgerald, O'Connor, & Wright, 2006*). Thus, it is possible that the differences in the auditory features of the stimuli (“beeps” and auditory speech) presented in Chapter 5 prevented generalization of perceptual learning. Furthermore, it is also possible that, like the differences between the tasks in the study by *Mossbridge et al. (2006)*, that simultaneity judgments for static versus dynamic stimuli differ significantly to prevent generalization following multisensory perceptual training.

Similarly, the visual system also has a limited capacity for generalization as described in Chapter 1. While much of the specificity described for visual perceptual learning is hypothesized to be due to changes in sensory processing following perceptual learning, it is also possible that some cognitive elements of visual perceptual learning may not transfer. Specificity for particular

reference-frames suggests that visual perceptual learning may occur at higher order levels. When perceptual training with sensory stimuli (e.g. visual orientation of the stimulus) remains constant but the reference-frame is changed (e.g. participants asked to report the stimulus with an orientation that is closest to vertical rather than horizontal axis), transfer of earlier perceptual learning does not occur, suggestive that perceptual learning does not only occur at the sensory level (Mastropasqua & Turatto, 2015). Thus, if the changes in temporal acuity we observe in Chapters 5 and 6 for audiovisual speech do occur at a cognitive level, it is possible that these changes are specific to the stimulus or at least the level of stimulus complexity.

Perceptual learning observed in Chapter 5 may also occur for different stimulus features, as the “flash-beep” stimulus is static while the audiovisual speech stimulus has numerous unique features, including dynamic visual motion cues and modulation of the auditory signal. If perceptual training does indeed enhance the perception of the motion cues present in visual speech, it would be interesting to investigate whether the changes in temporal acuity would transfer to other dynamic stimuli. Furthermore, if the enhancement of temporal acuity does result from enhanced processing of visual motion, it would be interesting to determine if either global motion or local motion features are important for changes in temporal acuity. The relative importance of focal elements of audiovisual speech (i.e. mouth movements) and more global features for multisensory integration and speech perception is unclear (V van Wassenhove, 2013).

It is possible that the lack of transfer between stimuli of different complexities that we observed following perceptual training to enhance perceptual acuity does not occur because the design of our training paradigm promotes specificity of perceptual learning. For visual perceptual learning, less extensive perceptual training is reported to facilitate better transfer of visual perceptual learning to other locations and stimulus orientations (Jeter, Doshier, Liu, & Lu, 2010). Thus, the initial rapid improvement observed for visual perceptual learning is sufficient to enable transfer of perceptual learning while more extensive and gradual perceptual learning

impairs transfer of perceptual learning. In our training paradigm, we observe a rapid change in temporal acuity following the first day of perceptual training and subsequently little to gradual change in temporal acuity over the following three days of perceptual training. Thus, in these subsequent training sessions (and possibly even during the later training blocks of the first day's training session), the changes in temporal acuity may become more specific to the trained stimulus, preventing transfer of changes in temporal acuity to other stimuli.

For visual perceptual learning, transfer (in terms of retinal location) also appears to be dependent on the actual precision (how demanding the stimulus presented is) of the stimuli presented during the transfer period rather than the precision of the stimuli during training (Jeter, Doshier, Petrov, & Lu, 2009). Thus, it is interesting to speculate whether the training procedures utilized in the majority of studies in the previous chapters involved training in a high precision environment (i.e. the training SOAs) and later testing (i.e. the post-training SJ test) in a low precision environment (i.e. the post-training SOAs). For participants in the studies described in Chapters 2, 3, 5 and 6, the SOAs utilized for perceptual training ( $\pm 150$ ,  $\pm 100$ ,  $\pm 50$ , and 0 ms) may have been a very precise stimulus in comparison to the SOAs utilized during the pre- and post-training SJ tests (generally SOA from +400 to -400 ms). Yet, evidence from the study in Chapter 3 suggests that training with SOAs of lower precision also does not facilitate perceptual learning. Thus, it may not be possible to facilitate generalization of enhancements of multisensory temporal acuity using the training paradigm described by these chapters.

#### *Perceptual Training that Enhances Temporal Acuity for Multisensory Speech is Less Durable*

The findings in Chapter 5 report that one week after the last perceptual training session, the TBW for audiovisual speech is not significantly narrower than the TBW measured prior to perceptual training. This suggests that the changes in temporal acuity for audiovisual speech following perceptual training may not be as durable as the changes in temporal acuity for “flash-beep” stimuli following perceptual training. One possible explanation is that perceptual learning

does not consolidate for speech stimuli due to immediate interference effects. Unlike “flash-beep” stimuli, immediate exposure to speech (i.e. talking to people) may negate some of the learning that occurs during the perceptual training session. It is possible that some of the aspects of perceptual learning that are observed in Chapter 5 for audiovisual speech are specific to the speech stimulus. Alternatively, the variability in the temporal relationships of speech may sufficiently differ from speaker to speaker to disrupt learning that occurred during the perceptual training session. Evidence suggests that visual perceptual learning may be disrupted by immediately performing a subsequent task for similar but different stimulus feature following initial perceptual training (Seitz et al., 2005). Conversely, a delay of 1 hour between the first and second perceptual learning tasks is sufficient for consolidation of the feature that is learned in the first task (Seitz et al., 2005). Thus, it is possible that in Chapter 5 there is insufficient time for consolidation to occur between perceptual training and additional speech exposure. The “flash-beep” stimulus, although a simple sensory pair, may be specific enough that similar stimuli are not likely to be encountered in a typical environment, or at least to the extent that speech is encountered. There is a slight trend, however, for the TBW for “flash-beep” stimuli to be wider at the pre-training SJ assessment (after Day 1) which was also observed on some days by Powers et al. (2009), suggestive not all of the increase in temporal acuity is retained for some period following perceptual training (Powers et al., 2009). Evidence from visual perceptual learning also suggests that extensive training with feedback is necessary for sufficient consolidation of perceptual learning that is not vulnerable to deterioration when exposed to additional stimuli (Dobres & Watanabe, 2012). While training with feedback did not seem to be sufficient enough to promote consolidation of the changes in temporal acuity for audiovisual speech, it remains possible that more extensive training over the course of more training sessions may have made the changes in temporal acuity for audiovisual speech more durable.

## Future Directions

The following will describe some aspects to consider for future explorations of multisensory perceptual learning.

### *Design of Multisensory Perceptual Training Paradigm*

One of the disadvantages of the utilizing a 2-AFC (yes/no) SJ task to index temporal acuity may be that the changes that develop following perceptual training result from a change in cognitive criterion and thus not represent changes in perceptual representations. Perceptual training tasks that employ a 2-IFC SJ task may avoid possible effects of a criterion for simultaneity (Powers et al., 2009). Many studies have thus compared changes in temporal acuity following perceptual training using both a 2-AFC and 2-IFC SJ tasks and found similar effects of perceptual training (Powers et al., 2009; Powers III et al., 2016). The 2-IFC task, however, is also subject to participant's response bias as participants may exhibit a preference for reporting one choice more often in one of the presentation intervals (García-Pérez & Alcalá-Quintana, 2011). Powers et al. (2009) reported a tendency for participants in the 2-IFC task to exhibit a response bias towards reporting the presentation of the visual and auditory stimuli on simultaneous catch trials in the second stimulus presentation interval as synchronous. This response bias was particularly strong for those participants for which perceptual training did not enhance temporal acuity (described by Powers et al. as static participants) (Powers et al., 2009). Thus, the 2-IFC may also not be ideal for perceptual training.

The advantage of using a 2-AFC SJ task is that it allows for a single judgment of the perception of simultaneity in comparison to the 2-IFC SJ task. It may, however, be valuable for future studies to consider the 3-AFC SJ for perceptual training. The 3-AFC variant of the SJ task is hypothesized to mitigate some of the cognitive biases that might affect the 2-AFC SJ task (Van Eijk, Kohlrausch, Juola, & van de Par, 2008). Unlike the 2-IFC SJ task, the 3-AFFC SJ would allow for participants to make a single sensory judgment.

Another consideration may be the manner in which the feedback signal is presented. For multisensory perceptual learning, it appears that a feedback signal may be necessary for perceptual learning to occur in the context of perceptual training. This is in contrast to visual perceptual learning may develop in the absence of feedback (Fahle & Edelman, 1993). In the studies described by the previous chapters as well as other studies of multisensory perceptual learning (Powers et al., 2009; Powers III et al., 2016), trial-by-trial feedback is provided during perceptual training sessions. Yet some have suggested that feedback may result in shifts in a cognitive criterion within a perceptual training session. For visual perceptual learning, block feedback is not typically reported to change criterion (Aberg & Herzog, 2012). Furthermore, some evidence suggests that block feedback may facilitate perceptual learning just as well as trial-by-trial feedback (Herzog & Fahle, 1997). Thus, it may be advantageous to consider the use of block feedback for perceptual training using a 2-AFC (yes/no) SJ task so as to minimize changes in criterion. Another possible strategy to avoid changes in criterion introduced by perceptual training with an SJ task is to measure perceptual learning on a subsequent day. This is supported by evidence from visual perceptual learning that changes in criterion observed with visual perceptual learning are observed to dissipate when perceptual learning is measured on a second day (Aberg & Herzog, 2012).

Although evidence from Chapter 2 suggests that multisensory perceptual learning occurs with multiple types of feedback (i.e. symbolic, social, and monetary feedback), there is some evidence that perceptual training may dependent upon the saliency of the feedback signal. For example, a decrease in participants' report of the SIFI is observed only after providing performance based monetary incentives (although the decrease in the report of the illusion appeared to not be durable as the SIFI was reported when feedback was removed at a similar level to when the SIFI was measured prior to training) (Rosenthal, Shimojo, & Shams, 2009). Thus, it may be important to at least consider the type of feedback provided during

perceptual training may affect the generalization of enhancements in temporal acuity or the durability of multisensory perceptual learning.

### *Potential for Perceptual Training to Enhance Temporal Acuity for those with ASD*

Atypical sensory function has been increasingly recognized as a feature of autism spectrum disorder (ASD) (Marco, Hinkley, Hill, & Nagarajan, 2011). Recently, a number of studies have reported atypical multisensory processing in ASD, describing decreased multisensory integration (Stevenson, Siemann, et al., 2014a, 2014b) and impaired temporal acuity (Foss-Feig et al., 2010), particularly for audiovisual speech stimuli (Stevenson, Siemann, Schneider, et al., 2014). Impairments of multisensory temporal acuity, in particular, have been hypothesized to be indicative of less reliable perceptual binding in ASD. Thus, for individuals with ASD a wide TBW may result in sensory cues being perceptually bound that should otherwise be discriminated, subsequently resulting in a less accurate perceptual representation of the sensory environment.

Perceptual training to enhance multisensory temporal acuity has been hypothesized as a strategy to ameliorate deficits of temporal acuity for those with ASD (Wallace & Stevenson, 2014). Importantly, as temporal acuity has been observed to correlate with the magnitude of multisensory integration (as reported by perception of the McGurk illusion) (Stevenson, Siemann, Schneider, et al., 2014), changes in temporal acuity may also parallel changes in more general multisensory functions and integration. Thus, it is interesting to speculate that if perceptual training is capable of enhancing temporal acuity for those with ASD, there may also be more generalized benefits of enhanced temporal acuity that results in better multisensory processing.

The results reported in Chapter 5 are of particular performance as this chapter reports that perceptual training is capable of enhancing temporal acuity for audiovisual speech (for typically developed adults). For individuals with ASD, deficits in temporal acuity appear to be

specific to audiovisual speech stimuli (Stevenson, Siemann, Schneider, et al., 2014). Thus, the findings in Chapter 5 suggest that perceptual training may be capable of enhancing temporal acuity for audiovisual speech in those with ASD. Additionally, the findings of Chapter 5 also suggest that perceptual training to enhance temporal acuity for multisensory stimuli, like “flash-beep” stimuli, may not elicit generalized improvements in temporal acuity that would result in enhanced temporal acuity for audiovisual speech. Future studies will be necessary to explore if perceptual training to enhance multisensory temporal acuity also generalizes to improve other aspects of multisensory perception, such as greater multisensory integration.

## References

- Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision, 12*(3). doi:10.1167/12.3.3
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput Biol, 5*(7), e1000436.
- Desantis, A., & Haggard, P. (2016). Action-outcome learning and prediction shape the window of simultaneity of audiovisual outcomes. *Cognition, 153*, 33-42.
- Dobres, J., & Watanabe, T. (2012). Response feedback triggers long-term consolidation of perceptual learning independently of performance gains. *Journal of Vision, 12*(8), 9.
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Research, 1242*, 136-150.
- Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision research, 33*(3), 397-412.
- Fister, J. K., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*.
- Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research, 203*(2), 381-389.  
doi:10.1007/s00221-010-2240-4
- García-Pérez, M. A., & Alcalá-Quintana, R. (2011). Interval bias in 2AFC detection tasks: sorting out the artifacts. *Attention, Perception, & Psychophysics, 73*(7), 2332-2352.
- Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage, 124*, 876-886.
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision research, 37*(15), 2133-2141.

- Jaekl, P., Pesquita, A., Alsius, A., Munhall, K., & Soto-Faraco, S. (2015). The contribution of dynamic visual cues to audiovisual speech perception. *Neuropsychologia*, *75*, 402-410.
- Jeter, P. E., Doshier, B. A., Liu, S.-H., & Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision research*, *50*(19), 1928-1940.
- Jeter, P. E., Doshier, B. A., Petrov, A., & Lu, Z.-L. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision*, *9*(3), 1-1.
- Kahnt, T., Grueschow, M., Speck, O., & Haynes, J.-D. (2011). Perceptual Learning and Decision-Making in Human Medial Frontal Cortex. *Neuron*, *70*(3), 549-559.  
doi:10.1016/j.neuron.2011.02.054
- Marco, E. J., Hinkley, L. B., Hill, S. S., & Nagarajan, S. S. (2011). Sensory processing in autism: a review of neurophysiologic findings. *Pediatric Research*, *69*, 48R-54R.
- Martin, J.-R., Kösem, A., & Van Wassenhove, V. (2015). Hysteresis in Audiovisual Synchrony Perception. *Plos One*, *10*(3), e0119365.
- Mastropasqua, T., & Turatto, M. (2015). Reference-frame specificity of perceptual learning: The effect of practice. *Vision research*, *106*, 1-6.
- Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., & Wright, B. A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. *The Journal of Neuroscience*, *26*(49), 12708-12716.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience*, *32*(18), 6263-6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, *29*(39), 12265-12274.
- Powers III, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports*, *6*, 23374.
- Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain topography*, *21*(3-4), 185-192.

- Seitz, A. R., Yamagishi, N., Werner, B., Goda, N., Kawato, M., & Watanabe, T. (2005). Task-specific disruption of perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(41), 14895-14900.
- Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & psychophysics*, *65*(2), 318-328.
- Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964-1973.
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*, *34*(3), 691-697.
- Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014a). Brief report: arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of autism and developmental disorders*, *44*(6), 1470-1477.
- Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014b). Evidence for diminished multisensory integration in autism spectrum disorders. *Journal of autism and developmental disorders*, *44*(12), 3161-3167.
- Tsilionis, E., & Vatakis, A. (2016). Multisensory binding: is the contribution of synchrony and semantic congruency obligatory? *Current Opinion in Behavioral Sciences*.
- Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific reports*, *5*.

- Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B-Biological Sciences*, 282(1804). doi:10.1098/rspb.2014.3083
- Van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & psychophysics*, 70(6), 955-968.
- Van Eijk, R. L., Kohlrausch, A., Juola, J. F., & Van De Par, S. (2010). Temporal order judgment criteria are affected by synchrony judgment sensitivity. *Attention, Perception, & Psychophysics*, 72(8), 2227-2235.
- van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain. *Frontiers in psychology*, 4, 388.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4), 1181-1186.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & psychophysics*, 69(5), 744-756.
- Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biol*, 4(10), e326.
- Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, 22(7), 1583-1596.
- Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *The Journal of Neuroscience*, 17(7), 2429-2444.

Wallace, M. T., & Stein, B. E. (2007). Early experience determines how the senses will interact.

*Journal of neurophysiology*, 97(1), 921-926.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding

window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64, 105-

123.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments.

*Experimental Brain Research*, 152(2), 198-210.