

BEHAVIORAL AND FMRI MEASURES OF CROSSMODAL PLASTICITY
INDUCED BY AUDITORY SENSORY SUBSTITUTION

Thesis by

Noelle R. B. Stiles

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2015

(Defended September 23, 2014)

© 2015

Noelle R. B. Stiles

All Rights Reserved

ACKNOWLEDGMENTS

I would like to thank my advisor, Shinsuke Shimojo, who has been a fantastic mentor throughout my years at Caltech, both pushing me toward excellence and encouraging a deep understanding of the scientific process. I would also like to thank my committee, Henry Lester, Doris Tsao, John O'Doherty, and Carmel Levitan, for their invaluable advice and guidance through the Ph.D. process. I would like to thank Vikram Chib for his patient and helpful fMRI training, advice, and data processing aid.

My collaborators inside and outside of Caltech have also dramatically improved my experience and allowed me to explore scientific questions that are not included in my thesis. In particular, Carmel Levitan, thank you for your scientific fellowship, and I look forward to continuing our work together. Luis Goncalves and Enrico Di Bernardo, thank you for enthusiasm and many excellent scientific discussions. Armand R. Tanguay, Jr., thank you for your collaboration, mentorship, and encouragement; you have added to my success in more ways than I could possibly list.

I also appreciate all of the undergraduate and graduate students who have worked on research projects with me at Caltech (Sepideh Farmani, Yuqian Zheng, Kijun Seo, Charlotte Yang, Juri Minxha, Daisy Lin, Kevin Li, Mason McGill, Elizabeth Lawler, and Allison Ban) and at USC (Ben McIntosh, Jennifer Crisp, and Janet Kim). Thank you for your enthusiasm and hard work.

Finally to my parents, thank you for your endless encouragement and support; I certainly would not be here without you.

Several organizations have supported my research and career at Caltech in many ways, for which I am extremely grateful. I appreciate a fellowship from the National Science Foundation Graduate Research Fellowship Program, the Mary Louise Remy Endowed Scholar Award from the Philanthropic Educational Organization, a Discovery Scholars Prize from the University of Southern California, and timely Travel Awards from the Caltech Graduate Office. I also appreciate research grant support from the Della Martin Fund for Discoveries in Mental Illness; the National Institutes of Health; and the Japan Science and Technology Agency, Core Research for Evolutional Science and Technology.

ABSTRACT

Sensory substitution (SS) aids the blind by encoding information from vision into a tactile or auditory stimulus. The major focus of this thesis is on the study of the crossmodal plasticity engendered through training on SS devices, and on the study and improvement of the rehabilitative potential of SS.

The effectiveness of SS as a rehabilitative device is unfortunately limited. Blind and blindfolded sighted individuals can be trained to interpret SS sounds visually and even learn object localization and identification, as well as depth perception. Despite this, participants require significant training (1 week to 3 months) to learn to use an SS device. Even after that they require significant attentional resources and top-down executive control to perform basic visual tasks. The laborious interpretation of SS is in stark contrast to the effortless visual interpretation even in complex and cluttered environments. Recognition, localization, constancies, and depth perception are attained in vision with ease and a surprising level of automaticity. Therefore, a major focus of this thesis will be studying whether SS devices can be made more intuitive and consequently more vision-like and useful to blind individuals. In particular, we have discovered that certain types of textural patterns provide intuitive associations between auditory and visual interpretation, suggesting both the utility of inherent crossmodal mappings for SS interpretation and a dramatic change in training paradigm. Surprisingly, we found that sounds derived from such textural patterns were correctly matched to images by naïve-sighted participants.

We also studied a crucial element of visual processing, constancies, with SS to determine whether SS users could acquire constancies via crossmodal plasticity and use them effectively. Further, while training the blind and sighted participants to learn constancies, we focused on improving the training procedures to make the device training more effective. Results showed that the sighted and blind participants could learn length and orientation constancy. We also found that spontaneous head-tilting movements while learning length constancy significantly correlated with improved task learning. Moreover, the improvement was transferred to a no-head-tilt condition. Overall, our results indicate that stimuli externalization, vision-like processing, and plastic sensory-motor integration are important and learnable elements of effective SS use.

The second focus of this thesis is the automaticity and topographic mapping of SS through crossmodal plasticity. Previously, auditory or tactile stimuli generated by sensory substitution have been shown to be processed in primary visual regions via crossmodal plasticity in blindfolded sighted and early or late blind individuals. Sensory substitution is therefore intrinsically crossmodal, and has a unique type of crossmodal plasticity between multimodal and unimodal cortical regions. Several studies have recorded this auditory-visual crossmodal interaction with sensory substitution, but very few have attempted to quantify the spatial, temporal, and attentional aspects of this plastic neural network. In this thesis, we will grapple with two of these elements: Spatial representation of sensory substitution stimuli in visual regions (*i.e.*, topographic mapping), and the role of attention (or automaticity) in crossmodal sensory substitution processing.

We used fMRI imaging to investigate the automatic nature of sensory substitution crossmodal plasticity. Sensory substitution has been shown to activate visual cortex with an auditory or tactile stimulus, in a similar pattern to visual processing of objects and locations. However, vision is also bottom-up, perceptual, and automatic. It is still unknown whether sensory substitution, like vision, can also be automatically processed by visual regions. The literature has assumed that automatic processing with SS is not possible, as basic functionality requires extensive training (1 week to 3 months). Unexpectedly, we show in this thesis that sensory substitution activates visual regions in a passive tasks, as well as in tasks distracting attention from the SS stimulus (with fMRI). These results indicate that SS interpretation and crossmodal plasticity is more perceptual, as opposed to cognitive or top-down-controlled, than previously believed.

The topographical mapping of visual space onto visual cortex via sensory substitution was also studied with fMRI imaging. Vision has a retinotopic map such that close regions on the retina are processed by neighboring regions in primary visual cortex. We used fMRI imaging to determine whether this type of mapping holds when the visual image is encoded into sound (with SS) and then activates visual cortex via crossmodal plasticity. Interestingly, we found that topographic mapping with SS is not entirely vision-like, but has similarities.

Overall, this thesis aims to investigate the crossmodal plasticity that enables SS interpretation, and to improve SS as a rehabilitative device for the blind. This thesis investigation shows that SS can be used to learn visual constancies, which are critical to visual rehabilitation of the blind and dependent upon plasticity. It also indicates that crossmodal mappings can be used to intuitively interpret SS with no encoding knowledge

and attentive efforts, thereby enabling training design recommendations that may shorten and improve training on SS. Finally, neural imaging studies therein investigated crossmodal plasticity of sensory substitution processing that make SS acquire aspects of vision's automaticity and retinotopic mapping.

TABLE OF CONTENTS

Acknowledgments	iii
Abstract	v
Table of Contents	ix
List of Figures and Tables	x
Chapter 1: Introduction	14
Chapter 2: Length and Orientation Constancy Learning with Auditory Sensory Substitution	56
Chapter 3: Crossmodal Intrinsic Mappings Make Auditory Sensory Substitution Effortless	88
Chapter 4: Automaticity of vOICe Crossmodal Plasticity	121
Chapter 5: Topographic Mapping of vOICe Crossmodal Plasticity	200
Chapter 6: General Discussion	234
Appendix A: Supplementary Data for Chapter 3	245
Appendix B: vOICe Training Procedures	249
Appendix C: Post-fMRI Scanning Questionnaire	259
Appendix D: Supplementary fMRI Data	262
References	273

LIST OF FIGURES AND TABLES

Figure 1.1. Concept web for thesis	17
Figure 1.2. Layout of thesis themes	18
Figure 1.3. Outline of Chapter 1	20
Figure 1.4. Schematic diagram of the vOICE device	24
Figure 1.5. Behavioral outcomes of sensory substitution training	36
Figure 1.6. Imaging with sensory substitution	40
Figure 1.7. rTMS with sensory substitution	43
Figure 1.8. Neural network with sensory substitution	46
Figure 2.01. Sighted orientation task data	67
Figure 2.02. Blind orientation task data	68
Figure 2.03. Sighted orientation task data divided by head tilt	69
Figure 2.04. Blind orientation task data divided by head tilt	70
Figure 2.05. Sighted length constancy data	73
Figure 2.06. Blind length constancy data	74
Figure 2.07. Length constancy head tilt and performance correlation	75
Figure 2.08. Vision and blind task improvement comparison	76
Figure 2.09. Orientation constancy task duration	79
Figure 2.10. Length constancy task duration	80
Figure 2.11. Head tilt and length constancy with a horizontal line	81
Figure 2.12. Head tilt and length constancy with a vertical line	83
Figure 3.01. Example of intuitive image and vOICE matching	93

Figure 3.02. Experiment design for visual-auditory matching	104
Figure 3.03. Experiment design for auditory distraction during visual-auditory matching	105
Figure 3.04. Experiment design for visual distraction during visual-auditory matching	106
Figure 3.05. Experimental design for tactile-visual matching	108
Figure 3.06. Select vOICe data and images	109
Figure 3.07. Tests of vOICe crossmodal mappings	111
Figure 3.08. Tests of vOICe crossmodal mappings summed across images	112
Figure 3.09. Auditory and visual attention distraction vOICe data	113
Figure 3.10. Correlation between bimodal matching data and edge metric	115
Figure 3.11. Data and images from matching of vOICe sound and tactile patterns	116
Figure 3.12. Correlation between the bimodal and unimodal tasks	118
Figure 4.01. Outline of Chapter 4	123
Figure 4.02. vOICe experiment layout	137
Figure 4.03. vOICe localization task setup	138
Figure 4.04. vOICe training flow chart	139
Figure 4.05. fMRI experiment diagram of the auditory pause detection task	140
Figure 4.06. fMRI experiment diagram of visual pause detection task	141
Figure 4.07. fMRI experiment diagram of the vOICe distract counting task	142
Figure 4.08. Sighted participant localization behavioral performance	150
Figure 4.09. vOICe localization behavioral performance in individual participants	151

Figure 4.10. Blind and sighted localization behavioral performance slope vs. intercept	152
Figure 4.11. Blind participant localization behavioral performance	154
Figure 4.12. fMRI data: Post – pre training vOICe noise sighted participants	159
Table 4.1. fMRI data: Post – pre training vOICe noise sighted participants	161
Figure 4.13. fMRI data: Post – pre training vOICe noise distract sighted participant	162
Table 4.2. fMRI data: Post – pre training vOICe noise distract sighted participants	164
Table 4.3. fMRI data: Post – pre training familiar sounds sighted participants	165
Table 4.4. fMRI data: Post – pre training blind and severe low vision participants	171
Figure 4.14. Post-experiment questionnaire results	183
Table 4.5. fMRI covariate data: Post – pre training vOICe noise distract sighted participants	185
Figure 4.15. fMRI data: Pre – post training vision noise pause detection task in sighted participants	190
Table 4.6. fMRI data: Sighted participants, vision noise pause detection [Pre – Post]	192
Table 4.7. fMRI covariate data: Pre – post training vision noise sighted participants	194
Figure 5.1. Diagram of the mapping of visual space to cortical activation	202
Figure 5.2. Retinotopic mapping of macaque striate cortex	203
Figure 5.3. fMRI experiment diagram of the vOICe localization task	209
Figure 5.4. fMRI experiment diagram of the vision localization task	210
Figure 5.5. fMRI imaging results: vOICe dot [Right – Left location] post-scan	

sighted participants	214
Table 5.1. fMRI imaging results: vOICe dot post scan sighted participants	216
Figure 5.6. fMRI imaging results: Vision dot [Right – Left location] post-scan sighted participants	218
Table 5.2. fMRI imaging results: Vision dot [Right – Left location] post-scan sighted participants	220
Figure 5.7. Post-experiment questionnaire results	222
Table 5.3. fMRI covariate data: vOICe dot [Right – Left location] post-scan sighted participants	223
Figure 5.8. fMRI imaging results: vOICe dot [Right – Left location] post-scan blind participants	226
Table 5.4. fMRI data: vOICe dot [Right – Left location] post-scan blind participants	227
Figure 6.1. Concept web for thesis	240
Figure 6.2. Layout of thesis themes	241

CHAPTER 1:
INTRODUCTION

Introduction

Sensory substitution (SS) encodes an image into a sound or tactile stimulation, and trained participants have been found not only to utilize the stimulus to coordinate adaptive behavior, but also to process it in early visual areas. Some superusers of a sensory substitution device have further claimed to subjectively experience a vision-like perception associated with device usage (Ward & Meijer, 2010). This chapter will not only go over the technical and historical perspective of SS, but will also more importantly highlight the implications of SS to blind rehabilitation and the potential of SS to reveal crossmodal perceptual organization.

Sensory substitution is processed like vision at cortical levels, but is transduced by audition (or somatosensation) at receptor levels; thus, it should be considered neither pure vision nor audition/somatosensation, but rather a third type of subjective sensation, or “qualia” (the absolute, first-person, quality of sensory experiences). If perceptual experience in sensory substitution is unique, do the same visual primitives hold? Are these visual primitives fundamental to all vision-like processing, or are they dependent on the visual sensory transduction process? Several other questions fundamental to the essential nature of visual experience also become feasible to investigate with this new broader definition of “visual” processing, such as holistic *vs.* local processing, static *vs.* dynamic recognition and depth perception, and perception based on purely sensory *vs.* sensory-motor neural processing. Studies with sensory substitution attempt to aid the

blind by understanding these questions and thereby improving both SS devices and the users' quality of life. Further, these investigations advance neuroscience by demonstrating the roles that neural plasticity and sensory integration play in the organization of visual perception. In short, SS provides scientists and philosophers with a new artificial dimension to examine perceptual organization processes.

Overview of Thesis

Sensory substitution studies are bifurcated along rehabilitation and basic neural science objectives. Sensory substitution can aid the blind by enhancing environmental perception, and navigation. Further, as previously mentioned, sensory substitution is a unique crossmodal recombination of modalities, and therefore provides an interesting perspective on crossmodal interactions at the neural and behavioral levels. This thesis serves both of these basic and applied science objectives; each chapter emphasizes one or the other of these aims, as illustrated in Figure 1.1.

The second and third chapters of this thesis use behavioral techniques to improve perception with sensory substitution. In particular, these chapters study making sensory substitution more perceptual and effortless to use, and training participants to learn constancies with the vOICE device, enabling object externalization and adaptive control of behavior. The second chapter discusses the training of sighted and blind participants to learn orientation and length constancy. A surprising outcome of this training is that dynamic interaction with the stimuli was critical to enhanced learning on the length constancy task. The third chapter focuses on shortening vOICE training and improving training outcomes by using innate crossmodal mappings. The results of Chapter 3 may

be used to intelligently incorporate crossmodal mappings into training, thereby re-focusing and enhancing it.

The fourth and fifth chapters of this thesis use fMRI techniques to understand the neural processing and plasticity that underlie sensory substitution learning. The fourth chapter focuses on determining whether attention is required for sensory substitution to activate the visual cortex in blind and sighted participants. The fifth chapter determines whether vOICE perception is contralaterally mapped from visual space to visual perception in the same way that vision is mapped.

This thesis begins to investigate some of the unknown features of sensory substitution. It investigates the similarities and differences between sensory substitution and visual perception. Several experiments within this thesis indicate that sensory substitution may be more automatic and perceptual than shown in previous studies with sensory substitution. Finally, the theme of the unique crossmodal nature of sensory substitution, and how to exploit it for rehabilitative gains, are highlighted in several experiments. Figure 1.2 illustrates several of these thesis themes and how they build toward improving blind participant rehabilitation. Further, Chapter 6 will discuss these thesis themes and their implications for sensory substitution in detail.

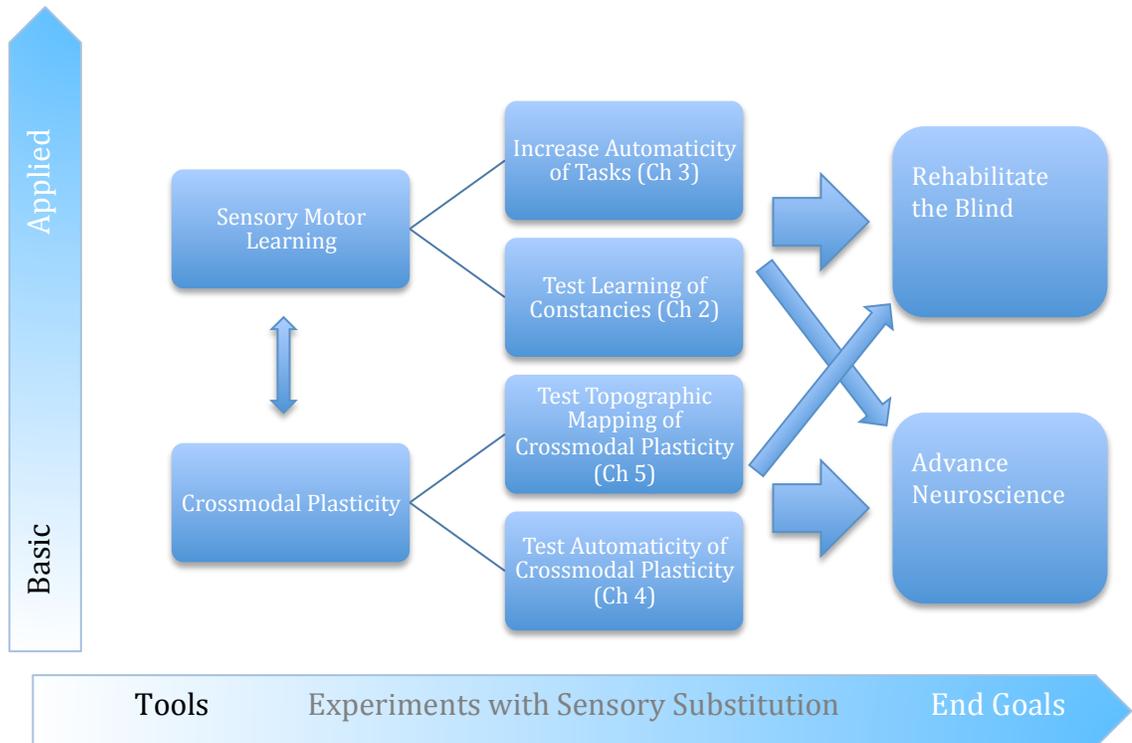


Figure 1.1. Concept web for thesis. This diagram spatially lays out the concepts developed in the thesis, and maps out several interesting inter-connections among concepts. In particular, it maps out the progress from tools to experiments to scientific goals for the thesis. It also shows the range from basic science to more applied science, and various cross-connections among the two. (Note: this figure is repeated as a review in Chapter 6.)

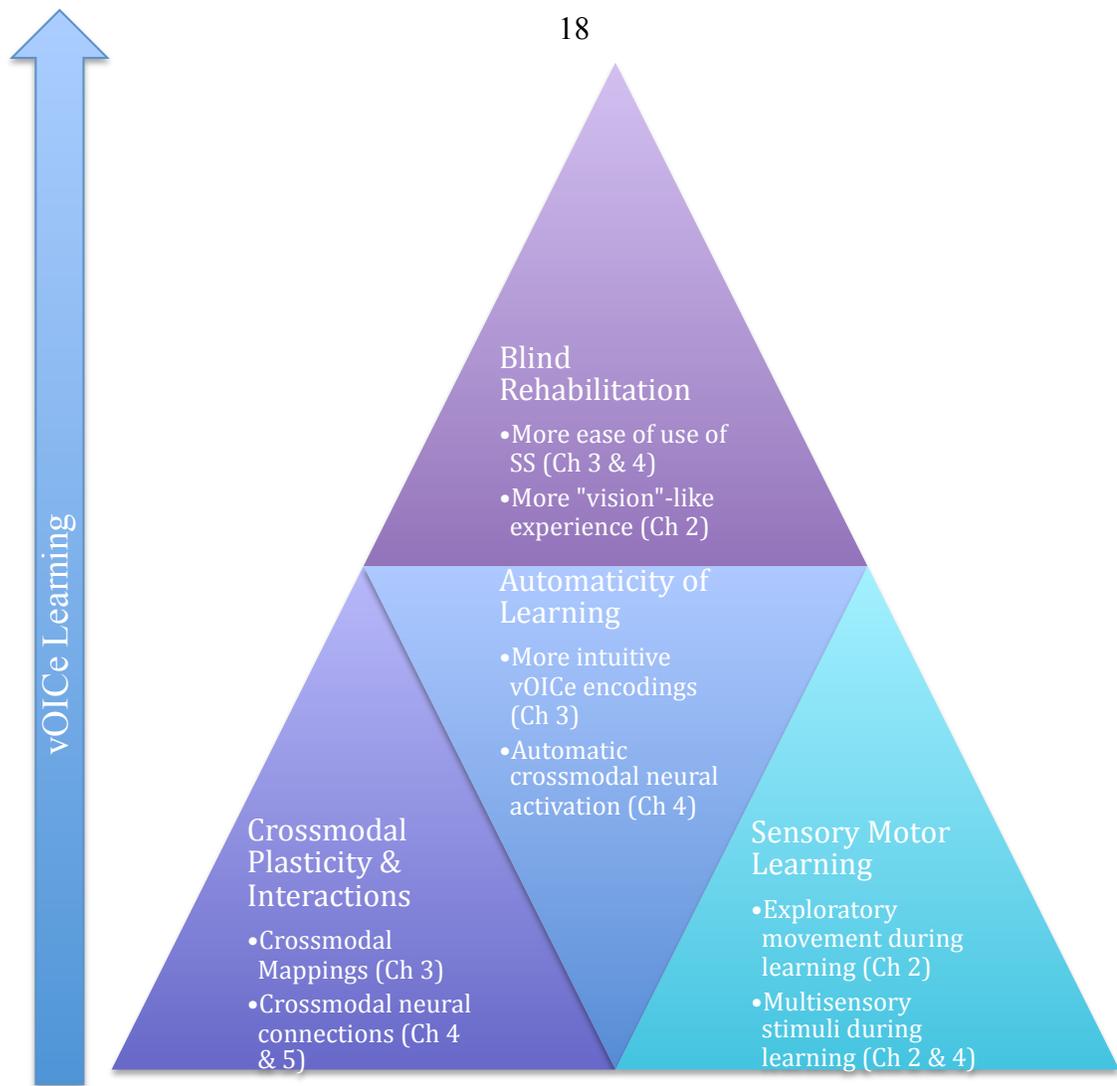


Figure 1.2. Layout of thesis themes. An alternative layout of thesis themes shows the crossmodal plasticity and sensory motor learning at the base of the pyramid, supporting the automaticity of perceptual processing and the rehabilitation of the blind. Each of the pyramid blocks has references to the chapters that relate strongly to those themes. (Note: this figure is repeated as a review in Chapter 6.)

Overview of Chapter

Chapter 1 is organized to include background information on sensory substitution and other relevant perceptual processes. To clarify the information covered in this chapter, a chapter outline is provided in Figure 1.3.

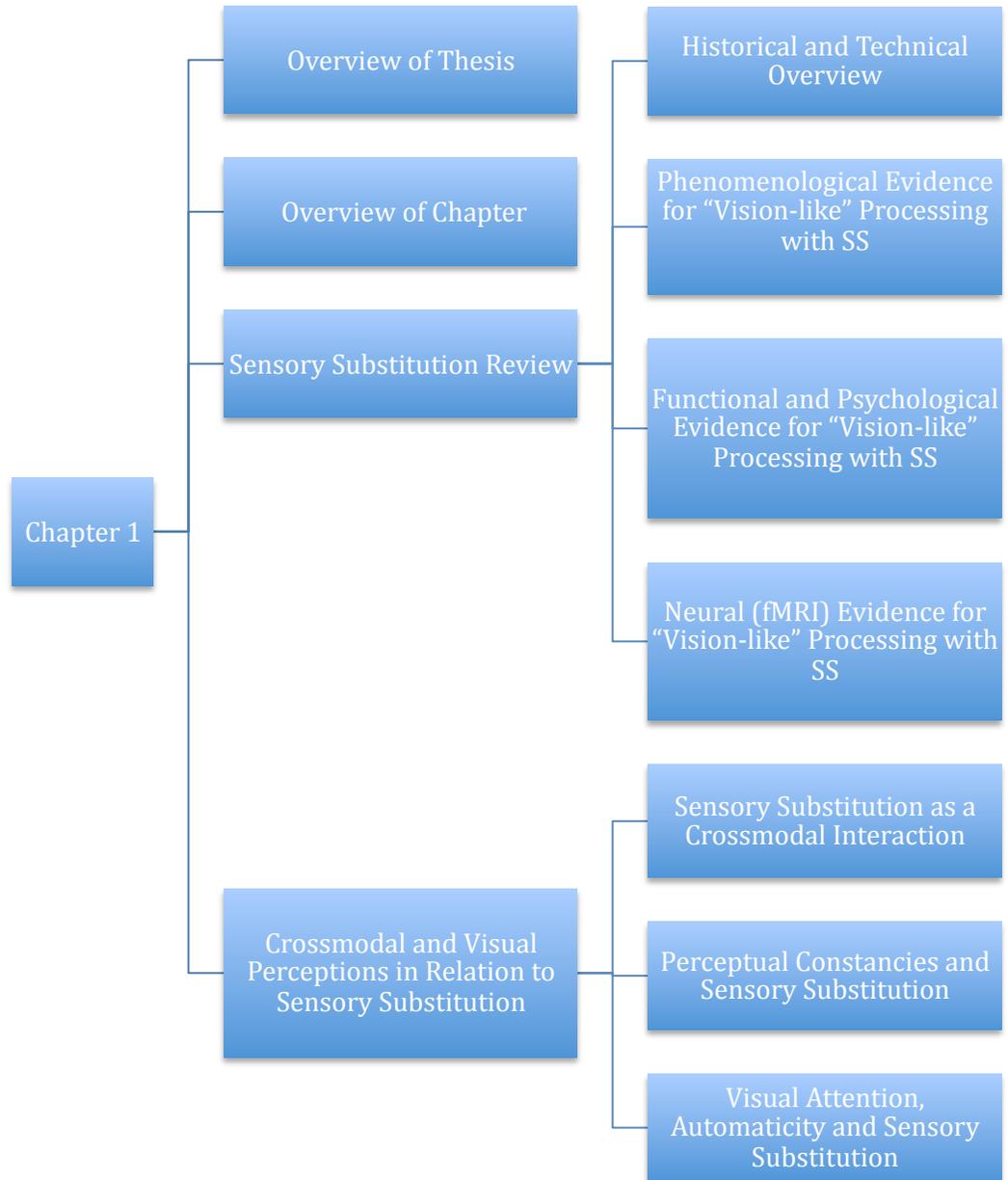


Figure 1.3. Outline of Chapter 1. This figure details the sections of Chapter 1 and their hierarchical structure.

Sensory Substitution Review

Historical and Technical Overview

Sensory substitution was designed as an aid to help the blind recover normal mobility and daily task functionality. Over 250 million people are visually impaired worldwide, with 39 million entirely blind (Visual Impairment and Blindness, 2012). The majority of the blind acquire blindness late in life (Resnikoff et al., 2004), but congenital blindness, or blindness inflicted near birth, still affects 1 out of every 3,300 children in developed countries (Bouvrie & Sinha, 2007). While specialized therapies, surgeries, and medication make most blindness preventable, blindness often cannot be ameliorated after the neural damage is complete. Therefore, several types of electronic prosthetic devices (such as retinal prostheses) have been designed that take over the function of the damaged neural circuitry by stimulating still-functional visual neurons (Humayun et al., 2003; Merabet, Rizzo, Amedi, Somers, & Pascual-Leone, 2005; Stiles et al., 2011; Winter, Cogan, & Rizzo, 2007). However, these devices are invasive, and are still in development. An alternative approach is *sensory substitution*, which encodes visual information into a signal perceived by another still-functional sensory modality, such as somatosensation of the skin or audition. Extensive crossmodal plasticity then enables the brain to interpret the tactile sensations and sounds visually.

Tactile sensation was first used by sensory substitution to transmit visual spatial information. The Tactile Visual Substitution System (TVSS) device used stimulators embedded in the back of a dental chair that were fed video by a camera mounted on a tripod (Bach-y-Rita, Collins, Saunders, White, & Scadden, 1969). With TVSS, six blind participants were anecdotally able to “discover visual concepts such as perspective,

shadows, shape distortion as a function of viewpoint, and apparent change in size as a function of distance” (Bach-y-Rita, et al., 1969). TVSS was later modified into the Brainport device that stimulates the tongue surface (Bach-y-Rita, Kaczmarek, Tyler, & Garcia-Lara, 1998) in order to reduce stimulation voltages and energy requirements as well as to utilize the high tactile resolution there.

Audition has also been used for sensory substitution with multiple types of encodings into sound. Early devices, such as the vOICe and PSVA devices, used a direct brightness-to-volume and pixel location to sound frequency transformation. The vOICe device encodes an image by representing vertical position as distinct frequencies, horizontal position as scan time (left to right), and the brightness of individual pixels as volume (Meijer, 1992) (Figure 1.4). The Prosthesis Substituting Vision by Audition (PSVA) device assigns a specific frequency to each pixel, and encodes brightness with volume (Arno et al., 2001; Capelle, Trullemans, Arno, & Veraart, 2002). More recent devices such as the Computer Aided System for Blind People (CASBlIP) and the Michigan Visual Sonification System (MVSS) have used 3D sound (encoded with head-related transfer functions) to encode the spatial location of objects (Araque et al., 2008; Clemons, Bao, Savarese, Austin, & Sharma, 2012).

Despite a diverse array of sensory substitution devices, none are currently commercially available or have a large user population. The limited commercial success of sensory substitution is likely due to the long duration (and substantial effort) required to learn a variety of basic visual tasks, and to the limited functionality realized once training is completed. Furthermore, a large part of the training improvement on psychophysical tests appears due to top-down executive control and concentration of

attention, even at the intermediate to advanced stages (Browne, 2003; Dunai, 2010; Ward & Wright, 2014). Recent devices, such as the MVSS and CASBlIP, hope to increase participant function and decrease training time by changing device encodings from vision-centric to audition-centric. By encoding spatial location in auditory coordinates, these devices exploit existing hardwired processing in auditory cortex while conveying useful information about obstacles. An alternative method to reducing training time and enhancing performance may be improvement of training methods, such as training that exploits intrinsic crossmodal correspondences (Pratt, 1930; Spence, 2011; Stevens & Marks, 1965) to make devices more intuitive, as will be explored in Chapter 3 of this thesis.

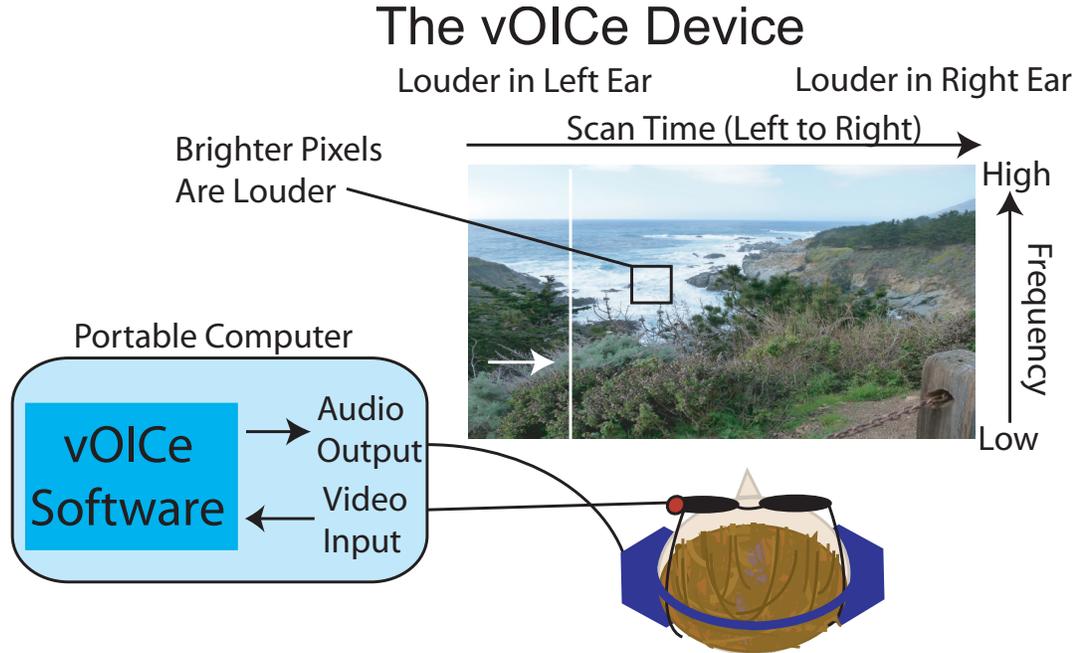


Figure 1.4. Schematic diagram of the vOICe device. A participant wears a pair of glasses with a camera attached that transmits live video to a portable computer. The computer runs the vOICe software, transforming the image into a soundscape by encoding the brightness of pixels into loudness of a sound frequency range that is high for upper pixels and progressively lower for middle and bottom pixels. This column of pixels is scanned across the image at 1 Hz with stereo panning (the scan rate is adjustable). The soundscape representing an image frame is communicated to the user via headphones.

Phenomenological Evidence for “Vision-like” Processing with SS

Sensory substitution generates activation in the primary visual cortex, but may also generate a vision-like perceptual experience, or have visual “qualia” (subjective, conscious quality of perception that can be verbally reported) in select long-term users. (Note that we only refer to the absolute unique quality of subjective perceptual experience here, regardless of whether the neural basis of qualia is a “hard problem” or not, as D. Chalmers has postulated (Chalmers, 1995).) In particular, late-blind vOICe user PF claims to have a visual experience with a sensory substitution device, and to even have color fill-in from previous visual experiences (Ward & Meijer, 2010). PF remembers colors in familiar items such as a strawberry, which she describes as a “red color with yellow seeds all around it and a green stalk”; whereas for unfamiliar objects, her brain “guesses” at the color such as “greyish black” for a sweater, and occasionally reduces the object detail to a line-drawing (Ward & Meijer, 2010). When rTMS was applied to her visual cortex, she claimed to have the visual experience damped, causing her to “carefully listen to the details of the soundscapes” instead of having an automatic “seeing” sensation, qualitatively linking visual activation to “visual” characteristics of the subjective experience (Merabet et al., 2009). The vOICe “visual” experience according to PF:

“Just sound?... No, it is by far more, it is sight!... When I am not wearing the vOICe, the light I perceive from a small slit in my left eye is a grey fog. When wearing the vOICe the image is light with all the little greys and blacks... The light generated is very white and clear, then it erodes down the scale of color to the dark black.” (Ward & Meijer, 2010)

Participant PF has not been the only blind user who has reported visual experiences with sensory substitution devices. A study with eighteen blind participants and ten sighted controls found that in the last three weeks of a three-month training period, seven blind participants claimed to perceive phosphenes while using a tactile sensory substitution device (Ortiz et al., 2011). Four out of seven participants with visual experiences retained light perception; they ranged in blindness onset from 1 year old to 35 years old. In most cases, the phosphenes appeared in the shape and angle of the line stimulus tactilely presented; the “visual” perception over time dominated the tactile perception (Ortiz, et al., 2011). The blind group with “visual” experience had activation in occipital lobe regions such as BA 17, 18, and 19 measured via electroencephalography (EEG); in contrast, the non-phosphene blind participants did not have visual activation (Ortiz, et al., 2011).

One critical aspect of the subjective visual experience is externalization, *i.e.*, the brain’s strong tendency to perceive visual inputs as external objects as opposed to something like visual images attached to the eyes (Palmer, 1999). Tactile devices have been studied for distal attribution of users (*i.e.*, the externalization of the stimulus) as defined by: 1. the coupling of participant movement and stimulation, 2. the presence of an external object, and 3. the existence of “perceptual space” (Auvray, Hanneton, Lenay, & O'Regan, 2005). Distal attribution was tested on 60 participants naïve to the auditory sensory substitution device and its encoding. Participants moved freely with headphones, webcam attached, and a luminous object in hand, and in some conditions were provided an object to occlude the luminous object. A link between participant’s actions and auditory stimulation was often perceived; this coupling perception occurred more often

than perception of distal object or environmental space. The coupling sensation between action and perception that participants perceived is perhaps another valuable aspect of the “qualia” of visual perception and sensory substitution. In fact, sensorimotor processing has been argued to be critical to visual awareness (O'Regan & Noe, 2001).

Key questions about “visual” sensations with sensory substitution remain. These include the connection between “visual” perception and functionality with the device, showing whether “visual” quality of experience enhances recognition and localization with sensory substitution. The neural mechanisms underlying visual perception with sensory substitution are also still unclear. Is “visual perception” via sensory substitution just mediated by primary visual areas, or do prefrontal and higher visual cortices play a key role? Further, a quantitative rTMS study of Ortiz’s participants that have “visual” experience may show whether the visual activation is necessary for their visual perception of sensory substitution stimuli. Deactivation of prefrontal regions (via rTMS) might demonstrate whether those regions are a part of a top-down cognitive network necessary to the distinctively unique subjective experience of “visual” nature with sensory substitution. These are feasible ideas to be tested in the future.

A major complication in visual activation and “visual” perception with sensory substitution is the role of visualization, or visual mental imagination, particularly in the late blind. The late blind have experienced vision and therefore are more familiar with visual principles, but also have the ability to activate visual cortex via visualization, or a mental effort to visually imagine a scene/object. PF is late blind (blindness onset at age of 21 years), and five out of seven of Ortiz’s blind participants with “visual” perception had blindness onset at the age of 4 years or later (Ortiz, et al., 2011). Therefore, it is

possible that the visual activation in these late-blind participants is due to top-down cognitive visualization rather than an automatic “visual” perception. The major evidence against mere visualization (as an alternative account) was limited to the qualitative claims that (1) the “visual” perception happens automatically, and (2) (in Ortiz’s participants) that tactile sensations fade and “visual” perception dominates. A quantitative study of the automaticity of “visual” perception with a sensory substitution device (*i.e.*, does it occur even when top-down attention is distracted) may further clarify the role of visualization in the sensory substitution “visual” experience. It will no doubt provide empirical seeds for theoretical re-consideration of the subjective aspects of perception, including the issue of “qualia.” Visualization as it relates to sensory substitution will be discussed in more detail in Chapter 4.

Functional and Psychological Evidence for “Vision-like” Processing with SS

In order for sensory substitution to be visual, it must also mimic the functional and psychological aspects of vision, or the organization and hierarchy of visual processing, that allow people to interact effectively with their environment. Key to visual functionality is depth perception with monocular depth cues such as perspective (parallel lines converge at infinity), relative size of objects, and motion parallax (lateral movement causes object movement to vary with distance) (Palmer, 1999). Furthermore, perceptual illusions are critical probes into vision-like processing, demonstrating the assumptions necessary to disambiguate a 3D world from 2D retinal images. Vision exhibits perceptual constancies that keep our perception of a given object the same despite varying observations of the environment, which may change the ambient brightness (brightness constancy), object distance (size constancy), color of illumination (color

constancy), tilt of the head (rotation constancy), and angle of the object (shape constancy), *etc.* (Palmer, 1999). Finally, effortless localization of objects in simple to cluttered environments and recognition of object properties and categories are critical to visual perception.

Recognition of artificial patterns and shapes has been investigated with tactile and auditory sensory substitution devices with positive results. Bach-y-Rita and colleagues tested five sighted participants on simple shape discrimination (such as circles and squares) with a Tongue Display Unit (a tactile sensory substitution device) (Bach-y-Rita, et al., 1998). Recognition performance averaged at 79.8 percent correct across shapes using arrays of 16, 25, 36, or 49 electrodes, and percent correct also improved with object size (Figure 1.5, A.a., line TO). Poirier *et al.*'s study tested pattern recognition with the PSVA (an auditory sensory substitution device) in blindfolded sighted participants (Poirier, De Volder, Tranduy, & Scheiber, 2007). Patterns were simple combinations of vertical and horizontal bars. Six sighted participants performed above 60% correct on element recognition before and after a training of 2 hours, and above 60% correct after training for pattern recognition (Figure 1.5, A.b.). Simple and complex pattern recognition was studied comparatively with auditory sensory substitution device PSVA in Poirier *et al.*'s behavioral analysis; they concluded that participants recognized the element size and spatial arrangement better than the pattern's element features (such as vertical bars and horizontal bars) (Poirier, Richard, Duy, & Veraart, 2006). Overall, sensory substitution studies show that users can recognize patterns and shapes when they are isolated on a plain background. However, recognition of one shape among many (as is most common in natural vision) has significantly less support.

Specialized object recognition has also been studied. In particular, sensory substitution face perception was investigated with PSVA (auditory sensory substitution device) for similar neural correlates to natural visual face perception, but participant recognition performance was not reported (Plaza et al., 2009).

Natural object recognition was tested in Auvray *et al.*'s 2007 study using the vOICe (auditory sensory substitution) (Auvray, Hanneton, & O Regan, 2007). Ten natural objects (such as a plant, shoe, and table) were identified by six sighted participants in an artificial white background (brightness was inverted before sonification) in an average of 42.4 seconds each (Auvray, et al., 2007). Participants listed 1.6 objects on average before choosing the correct object. The time to identification improved over training (from 57.6 seconds to 34.7 seconds), and varied among object type and individual participants. Categories of objects were studied with the 10 natural objects with 9 additional objects in the same category of an original object. Participants performed above chance at recognizing specific objects even within the same category, and participants were more accurate when there were fewer objects in each category.

A majority of the studies on object recognition with sensory substitution have focused on artificial stimuli in simplified environments. Thus far, no studies yet have explored natural objects in natural environments (such as finding a shirt in a closet, or a clock on a nightstand) or the role of distractor objects to object perception (such as recognizing a object in the center of the field of view, with two objects to the left and right). A potential reason is that artificial patterns are easier to identify, and also can be manipulated to test for sensory substitution resolution as well as quantify objects' complexity relatively easily, with a hope that more cluttered scenes would eventually

become recognizable in the progress of training. Several key visual questions, such as spatially segregating objects, object recognition independent of point of view (*i.e.*, shape constancy), and differentiation of shadows and reflections from physical objects, remain unanswered.

Vision is to perceive “what is where by looking” (Marr, 1982). Recognition studies investigated the “what” element of perception, and now, localization studies will highlight the “where” element of vision. Clinically, object localization has been most commonly studied with locomotion through a maze of obstacles. Chebat and his collaborators constructed a life-sized maze consisting of white hallway with black boxes, tubes, and bars horizontal (on the floor or partial protruding from the wall) or vertical (aligned with left or right wall)(Chebat, Schneider, Kupers, & Ptito, 2011). Sixteen congenitally blind and eleven sighted controls navigated the maze with a tactile display unit (10 × 10 pixels), and were scored for obstacle detection (pointing at obstacle) and obstacle avoidance (walk past the obstacle without touching it) (Figure 1.5 B.a.). Congenitally blind were able to detect and avoid obstacles significantly more accurately than the sighted controls. Both groups performed the tasks above chance. Larger obstacles were easier to avoid and detect than smaller obstacles, and step-around obstacles were easier to negotiate than step-over obstacles. Other localization studies have investigated artificial maze environments and tracking of stimuli in 2D and 3D space (Chekhchoukh, Vuillerme, & Glade, 2011; Kupers, Chebat, Madsen, Paulson, & Ptito, 2010).

Studies have also investigated localization via a pointing task, and the value to SS learning of SS device use in daily life. A study by Proulx and colleagues (2008) showed

that auditory sensory substitution localization was enhanced when participants were allowed to use the SS device in normal life (in addition to device assessments), compared to participants who only used the device during assessments (Proulx, Stoerig, Ludwig, & Knoll, 2008). Auvray and colleagues (2007) used an auditory sensory substitution device to study the accuracy of localization with a pointing task (Figure 1.5, B.b.) and found that 7.8 cm was the mean error for pointing at 4 cm diameter ball (Auvray, et al., 2007). The pointing inaccuracy varied proportionally with distance to the handheld camera (vertically aligned with the participant's elbow).

Depth perception and illusions are also a key part of visual processing. With sensory substitution's monocular camera and low resolution, it can be especially challenging for users to learn. Nevertheless, sighted users have been found to have key illusions of monocular depth perception, and other visual illusions. As described earlier in this chapter, Renier and colleagues have tested for perception of the Ponzo illusion with a auditory sensory substitution device, and found that blindfolded sighted participants could perceive it similarly to the sighted, but early-blind participants could not (Renier, Laloyaux, et al., 2005). Investigation of the vertical-horizontal illusion (vertical lines appear longer than horizontal lines) showed that sighted participants could perceive this illusion with an auditory sensory substitution device, but early blind participants could not perceive it (Renier, Bruyer, & De Volder, 2006). These results may indicate either that previous visual experience is essential for the perception of certain illusions, or that the duration of training may have been too short or superficial. Testing late-blind participants may further elucidate why congenitally blind participants did not perceive these illusions.

The perceptual organization of sensory substitution perception has many properties yet to be determined. Recognition and localization properties in natural environments are not thoroughly quantified; nor are performances in cluttered environments, or in shadowy and glare-ridden settings. Further questions as to what could be sensory substitution primitives (such as edges or spatial frequencies in vision) have not been answered. Scene perception with sensory substitution is also ambiguous. Questions such as: Can spatial relations of scene be generated with sensory substitution, and How much does it depend on past visual experience and the mode of stimulation (auditory or visual), are still unanswered. The active allocation of attention via gaze is also a critical component of the normal visual function that is entirely absent in sensory substitution encodings. Does the absence of active sensation inhibit the processing of sensory substitution stimuli and the generation of choice? Or instead, would exploration/orienting with the head turn compensate for the gaze shift easily with minimal training? How does the absence of the gaze cascade impact preference in the sensory substitution “visual” experience (Shimojo, Simion, Shimojo, & Scheier, 2003)? Finally, Gestalt binding principles of proximity and shared properties may or may not be perceived with sensory substitution, and may be controlled by the transducing modality (somatosensation or audition) or the processing modality (vision).

Neural (fMRI) Evidence for “Vision-like” Processing with SS

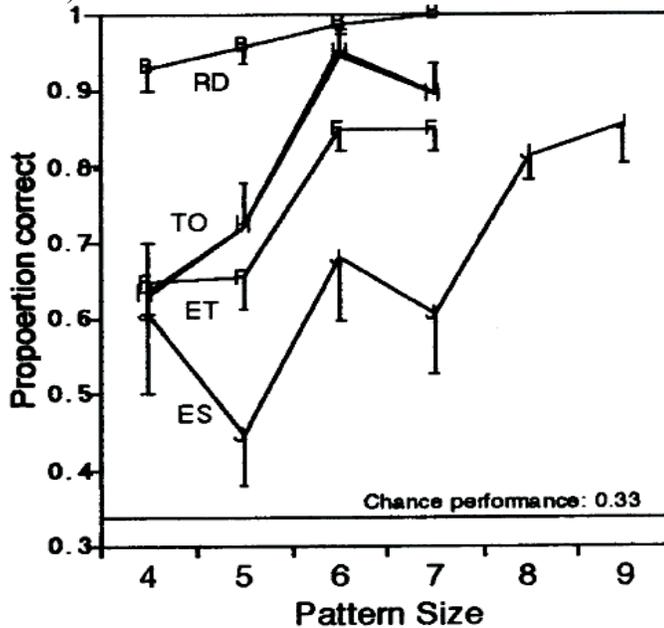
Neural imaging and stimulation studies have recently shown visual activation with limited SS device usage in sighted, late blind, and early blind participants. In 2007, Poirier *et al.* reviewed sensory substitution imaging studies, concluding that early blind users use primarily crossmodal plasticity and blindfolded sighted users mainly visual

imagery to generate visual activation with sensory substitution use (Poirier, De Volder, & Scheiber, 2007). PET and fMRI studies with tactile and auditory SS devices have shown activation in BA 17, BA 18, and BA 19 with recognition and localization tasks in early and late blind as well as occasionally blindfolded sighted participants (Amedi et al., 2007; Arno, et al., 2001; Kupers, et al., 2010; Merabet, et al., 2009; Poirier, De Volder, Tranduy, et al., 2007; Poirier, De Volder, & Scheiber, 2007; Poirier, De Volder, Tranduy, & Scheiber, 2006; Ptito, Moesgaard, Gjedde, & Kupers, 2005; Renier, Collignon, Poirier, Tranduy, Vanlierde, Bol, Veraart, & De Volder, 2005; Renier & De Volder, 2010; Renier, Laloyaux, et al., 2005). Early PET studies showed activation in occipital cortex for early blind participants, but not for sighted participants (Arno, et al., 2001; Ptito, et al., 2005). fMRI imaging studies later found visual activation with sensory substitution use in sighted participants with pattern recognition and localization, in particular in visual areas within the dorsal and ventral streams (Poirier, De Volder, Tranduy, et al., 2007; Poirier, De Volder, et al., 2006) (Figure 1.6, B).

Visual activation due to sensory substitution use has also been shown to be functionally correlated to the task performed during scanning. Amedi and colleagues showed with fMRI imaging that the lateral occipital tactile-visual (LOtv) area known to interpret object shape was also activated by auditory sensory substitution device usage during a shape task (Amedi, et al., 2007) (Figure 1.6, A). Plaza and collaborators in 2009 demonstrated that PSVA could activate the fusiform face area with face stimuli in blindfolded volunteers (Plaza, et al., 2009). Renier *et al.* investigated depth perception with a SS device, and found that blindfolded sighted participants could perceive the Ponzo illusion and had activation in occipito-parietal cortex while exploring 3D images

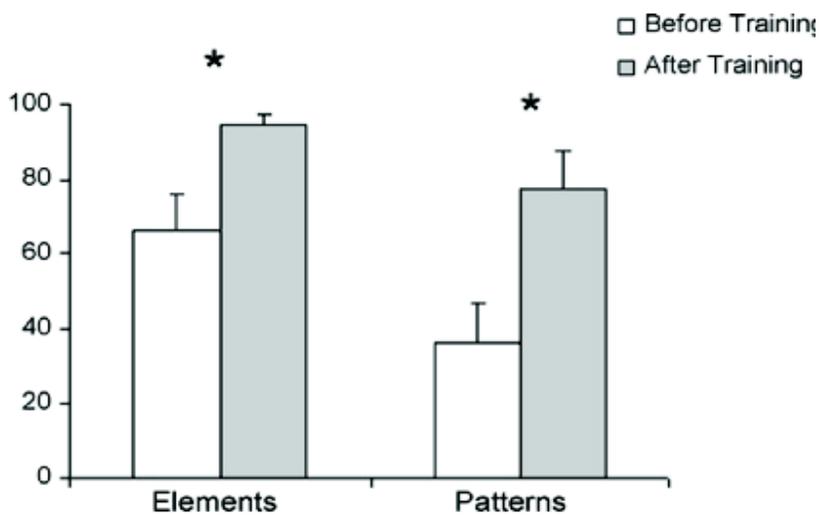
with PET imaging (Renier, Collignon, Poirier, Tranduy, Vanlierde, Bol, Veraart, & De Volder, 2005; Renier, Laloyaux, et al., 2005). Plaza and collaborators in 2012 compared neural activation for an orientation and localization task using images encoded into sound or presented visually (Plaza, Cuevas, Grandin, De Volder, & Renier, 2012). Sighted participants' neural activation was stronger in the right superior parietal lobule (BA 7) for the localization task in comparison to the orientation task in both auditory sensory substitution and vision.

Figure 1.5.A.a. Pattern Recognition, Tactile Sensory Substitution (Bach-y-Rita, et al., 1998)



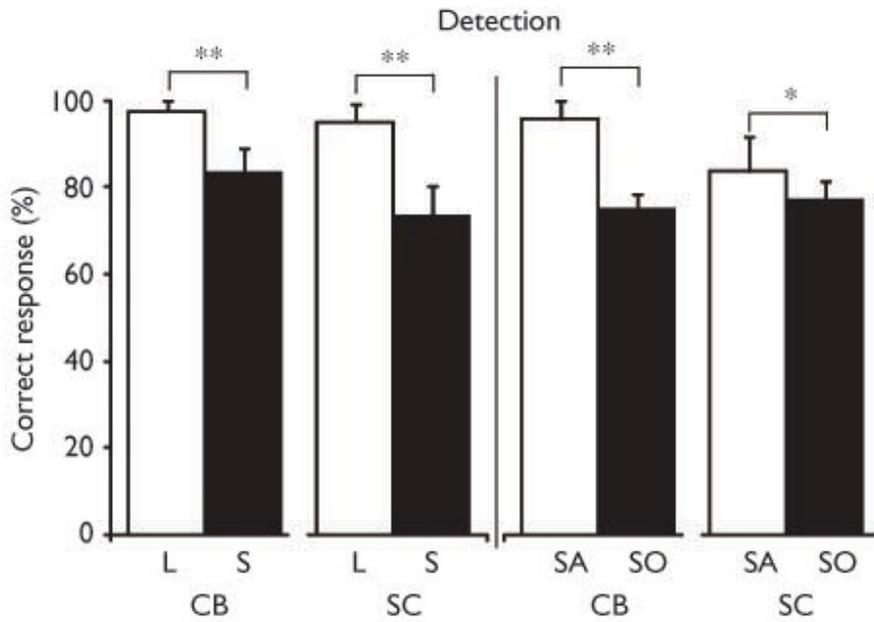
RD: Fingertip perceived raised dots,
 TO: Electrotactile tongue discrimination
 ET: Fingertip electrotactile discrimination
 (participant dynamically modulate current),
 ES: Fingertip electrostatic stimulation

Figure 1.5.A.b. Pattern Recognition, Auditory Sensory Substitution (Poirier, De Volder, Tranduy, et al., 2007)



* Statistically significant difference between before and after training (Elements: Wilcoxon test for paired samples: $Z = 1.99$, $p < 0.05$; Patterns: Wilcoxon test for paired samples: $Z = -2.23$, $p < 0.03$)

Figure 1.5.B.a. Object Localization, Tactile Sensory Substitution (Chebat, et al., 2011)



CB: Congenitally Blind, SC: Sighted Controls,
 L: Large Object, S: Small Object,
 SA: Step-Around Obstacle, SO: Step-Over Obstacle
 (* $p \leq 0.05$; ** $p \leq 0.001$)

Figure 1.5.B.b. Object Localization, Auditory Sensory Substitution (Auvray, et al., 2007)

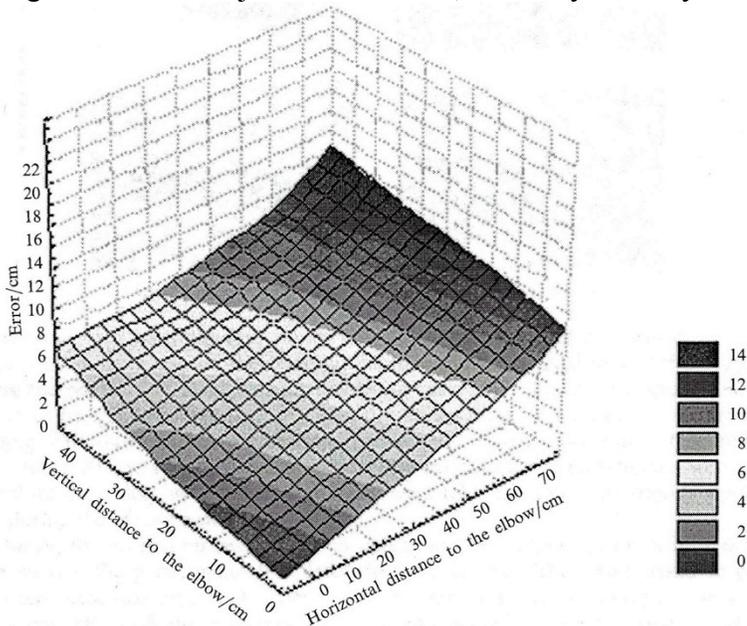
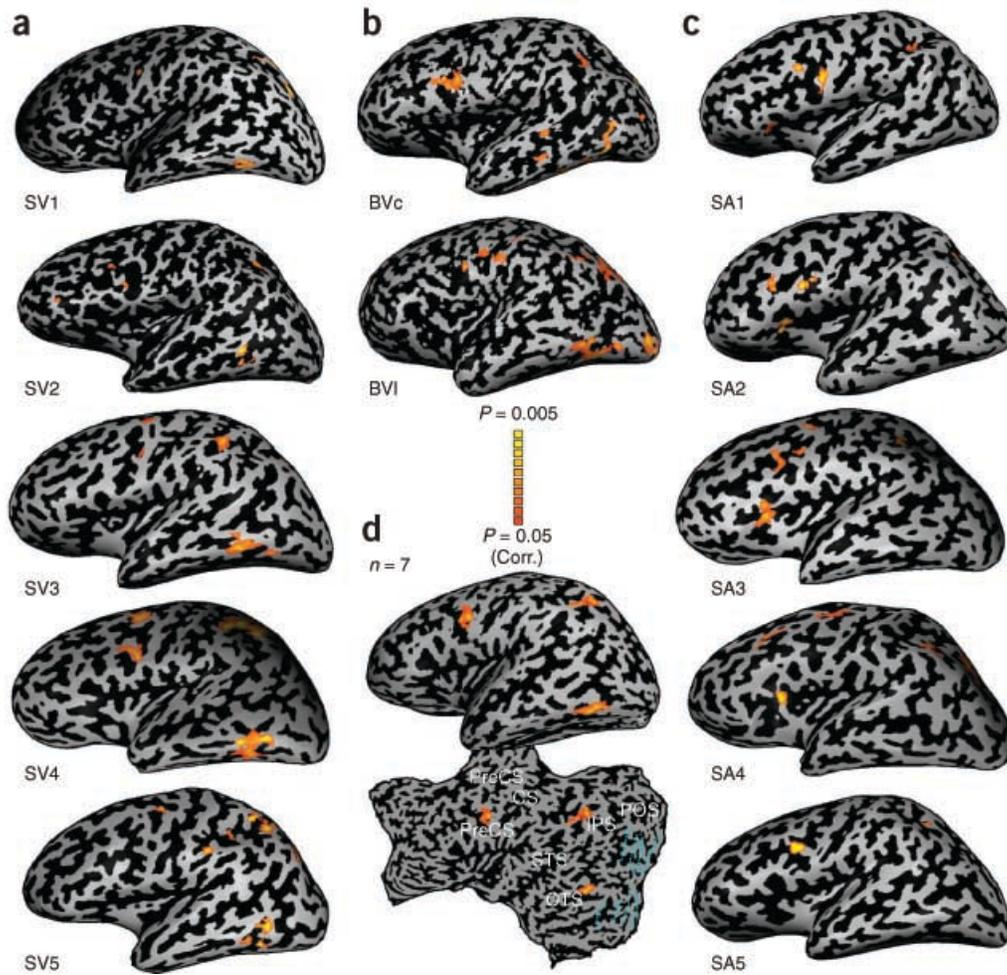


Figure 1.5. Behavioral outcomes of sensory substitution training. Psychophysical testing with tactile and auditory sensory substitution devices has had similar outcomes. Object recognition testing with the Tongue Display Unit (A.a.) has shown a correlation between the pattern size and the proportion correct; all participants exceeded the chance performance. Pattern recognition with an auditory device (A.b.) significantly improved with training, and had a similar average percent correct as tactile pattern recognition (between 0.6 and 0.8 proportion correct). Obstacle localization in an uncluttered maze environment with a tactile device (B.a.) had between 0.8 and 1 proportion correct for most object types. Localization of a 4 cm diameter ball with an auditory device showed that inaccuracy increased with distance to the object (webcam to view environment was held in the right hand and aligned with the elbow). (Auvray, et al., 2007; Bach-y-Rita, et al., 1998; Chebat, et al., 2011; Poirier, De Volder, Tranduy, et al., 2007)

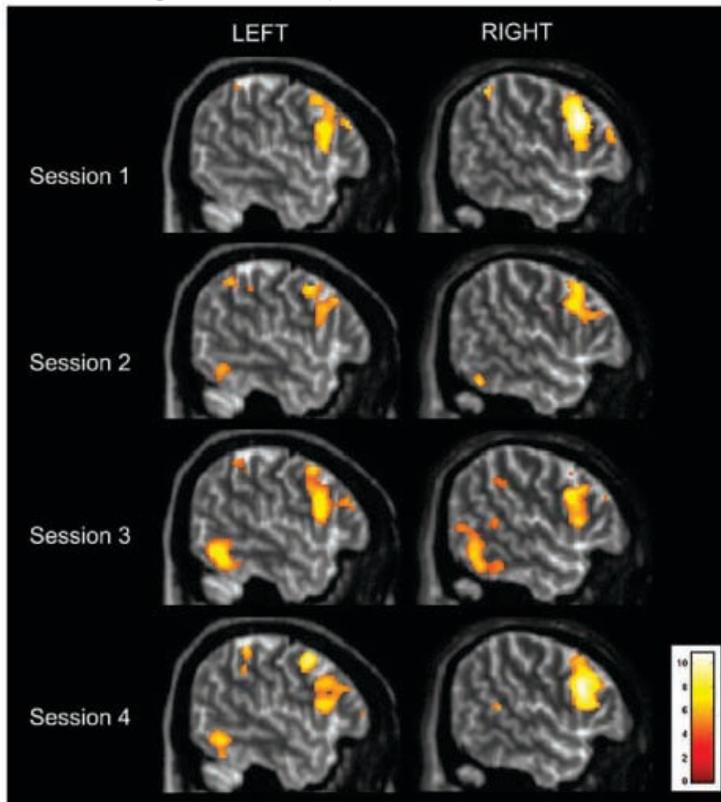
Even non-sensory-substitution binding of crossmodal stimuli can generate visual activation from unimodal stimuli. Zangenehpour and Zatorre found that training on the spatial and temporal congruence of beeps and flashes activated visual cortex even in the auditory-only condition (Zangenehpour & Zatorre, 2010). Therefore, visual cortex can be trained to respond to audition if the participants are taught to associate temporally and spatially co-located beeps and flashes. This indicates that a critical part of training-induced plasticity is simultaneous stimulation of sensory substitution (audition or somatosensation) and vision (for sighted participants), potentially due to Hebbian learning. Hebbian learning can also be potentially extended to the blind if stimuli are felt by the hand simultaneously with stimulation by sensory substitution.

Figure 1.6A. Activation in Blind and Sighted with a Shape Estimation Task (Amedi, et al., 2007)



a. Single sighted participant's neural activation, b. Blind participant neural activation, c. Single sighted participant activation from auditory control task, d. Average across seven vOICE trained users (participants in a and b).

Figure 1.6B. Sighted Participant Activation as a Function of Training Session on a Pattern Recognition Task (Poirier, De Volder, et al., 2006)

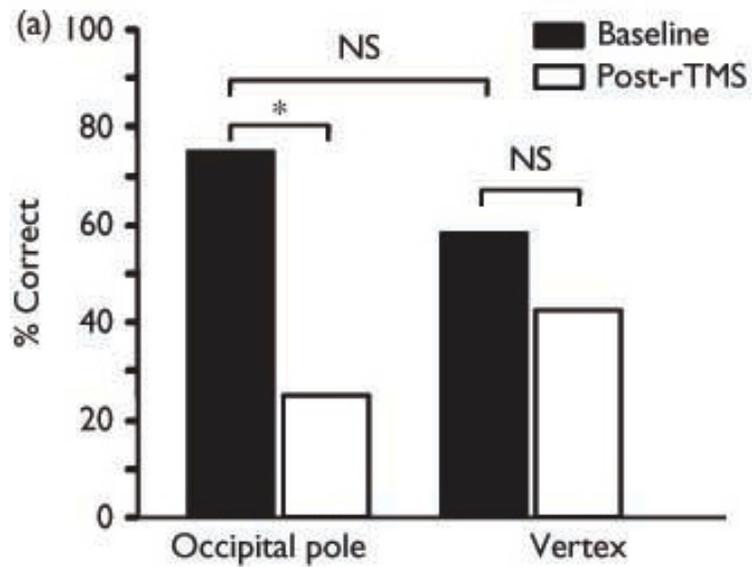


Voxels corrected for multiple comparisons in the whole brain and threshold exceeding $p < 0.05$. Six sighted participants.

Figure 1.6. Imaging with sensory substitution. Neural activation was shown on the left occipito-temporal cortex in all sighted and blind expert users during sensory substitution shape classification (A.a. – A.b.), whereas sighted users did not have visual activation with auditory control task (A.c.). Averaged results show activation in several multimodal regions (A.d.). During a sensory substitution pattern recognition task, six sighted participants showed a progressive increase in occipital activation with training on an auditory sensory substitution device (B.) (Amedi, et al., 2007; Poirier, De Volder, et al., 2006).

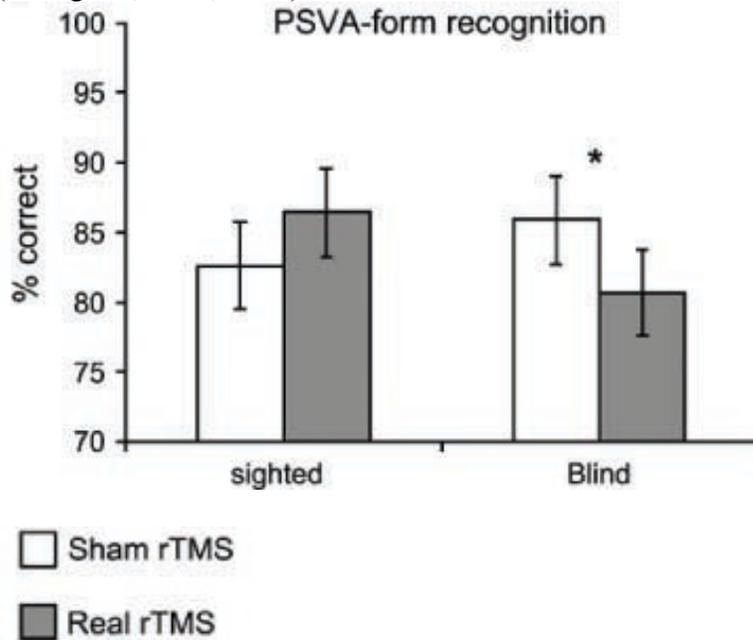
fMRI and PET studies have demonstrated that visual cortex activation correlates with sensory substitution use, but cannot prove causality. Repetitive Transcranial Magnetic Stimulation (rTMS) deactivates a region of cortex, examining the possible causal link between neural activation and participant performance. Collignon and colleagues applied rTMS to the right dorsal extrastriate occipital cortex of seven sighted and seven early blind participants (both trained on the PSVA auditory sensory substitution device) preceding sensory substitution pattern recognition (Collignon, Lassonde, Lepore, Bastien, & Veraart, 2007). Early blind participants had longer reaction times and lower accuracies with rTMS applied as compared to a sham rTMS condition; sighted participants had no performance change (Collignon, et al., 2007) (Figure 1.7, B). Merabet *et al.* also deactivated with rTMS occipital peristriate regions of a late blind sensory substitution superuser, PF, and demonstrated a decrement in recognition accuracy relative to pre-rTMS and post-sham rTMS conditions (Merabet, et al., 2009) (Figure 1.7, A). In the tactile domain, TMS applied to occipital cortex elicited somatotopic tactile sensations in blind but not blindfolded sighted users of a tactile sensory substitution device (Kupers et al., 2006). Overall, rTMS studies indicate that the blind users of sensory substitution devices functionally and causally recruit the occipital cortex, potentially due to long-term crossmodal plasticity from visual deprivation.

Figure 1.7A. rTMS on a Late Blind Auditory Sensory Substitution Expert (Merabet, et al., 2009)



NS: Not Significant, *: $p < 0.05$

Figure 1.7B. rTMS on Early Blind Auditory Sensory Substitution Users (Collignon, et al., 2007)

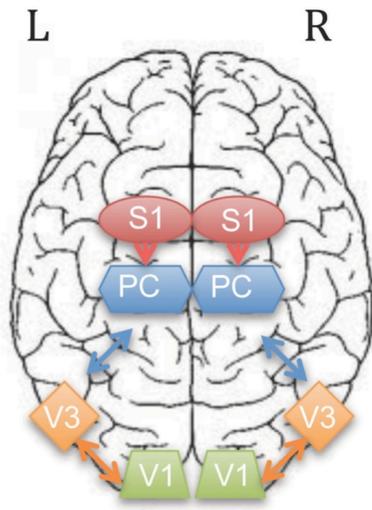


*: $p < 0.05$, Error bars indicate standard errors.

Figure 1.7. rTMS with sensory substitution. Repetitive Transcranial Magnetic Stimulation (rTMS) decreases neural activation and influences behavior, thereby generating a causal link between behavioral outcomes and neural region activation. rTMS of an occipital region significantly reduced percent correct at object identification in an expert vOICE user, PF (A.). PF's recognition was not significantly impaired by rTMS of a vertex location. Seven early blind participants were also impaired at the sensory substitution pattern recognition task with rTMS to right dorsal extrastriate occipital cortex (B.). Seven sighted participants' performance was not significantly affected by rTMS (B.) (Collignon, et al., 2007; Merabet, et al., 2009).

Dynamic Causal Modeling (DCM) studies in the blind have constructed a crossmodal network for auditory and somatosensory connections to the visual cortex (Fujii, Tanabe, Kochiyama, & Sadato, 2009; Klinge, Eippert, Roder, & Buchel, 2010). It remains to be shown whether these networks are used in blind participants with sensory substitution, and whether the crossmodal network in the sighted is similar to, or different from blind participants. Nevertheless, literature on functional connectivity of sensory substitution “stimuli” and dynamic causal modeling of the blind can be used to generate several neural network possibilities (Figure 1.8, A and B) with feedforward and feedback connections. The network likely includes the primary sensory region of the transducing modality (somatosensation or audition), which connects to a multimodal region that further connects to primary visual regions (V3, V2, or V1). The filtering of stimuli as sensory substitution stimuli or natural stimuli could occur at the primary region of transducing modality (A1 or S1) or the multimodal region. More studies on the specificity of the plasticity would be required to elucidate this. The role of prefrontal regions in top-down cognitive processing of the crossmodal stimulus has yet to be shown. More critically, which specific regions in the network are casually linked to performance, and therefore the roles that the regions play in stimulus processing, have yet to be fully determined. Feedback between visual regions and the multimodal regions may play a significant role in stimulus processing, yet the degree of feedback in sensory substitution processing is unclear. Motor regions and other primary sensory regions may also play an important role in plastic changes in the sensory substitution neural network.

A. Tactile Sensory Substitution Neural Network



B. Auditory Sensory Substitution Neural Network

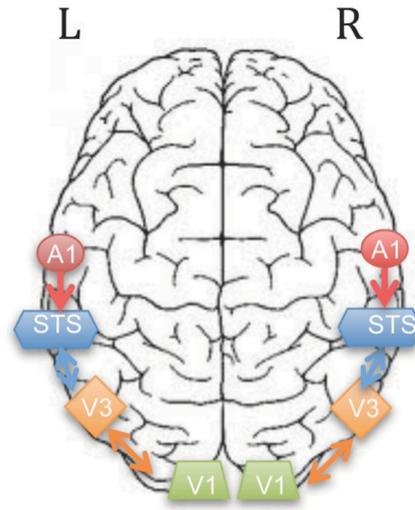


Figure 1.8. Neural network with sensory substitution. Visual, auditory, and tactile regions generate a neural network in blind and sighted sensory substitution users that processes sensory information within a feedforward and feedback hierarchy (A. for tactile devices and B. for auditory devices) (Poirier, De Volder, & Scheiber, 2007). The sensory information is first filtered by primary sensory regions (A1 or S1 for auditory and tactile devices, respectively). Sensory information is then communicated to multimodal regions (such as STS or Parietal Cortex [PC]) and forwarded to primary visual regions (V3, V2 [not shown] or V1). It is also likely that feedback and reiterative processing play a role in the perception of the sensory substitution stimuli.

Structural connectivity between and within sensory regions has also been measured in the blind and sighted using MRI Diffusion Tensor Imaging (DTI) and Diffusion Tensor Tractography (DTT). Shimony *et al.* found that early blind individuals maintained their white matter tracts between visual cortex and orbital frontal and temporal cortices (Shimony et al., 2006). Shu *et al.* used DTT and found that early blind participants had reduced connectivity compared to sighted controls (Shu et al., 2009). Overall, white matter neural imaging indicates interesting similarities and differences between the sighted and blind populations.

Crossmodal and Visual Perceptions in Relation to Sensory Substitution

Sensory Substitution as a Crossmodal Interaction

Regardless of the specific encoding employed, sensory substitution is intrinsically crossmodal, as the information from the transducing modality is communicated to visual cortex for processing by means of either intrinsic mapping (such as matching of high pitch and high spatial location) or neural plasticity engendered through training. The crossmodal interactions utilized by sensory substitution exist as both hardwired developmental connections and plasticity-induced changes in adulthood.

As an example of a more hardwired crossmodal interaction, the Illusory Flash or Double Flash Illusion (in which a single flash accompanied by two short sounds is perceived to be doubled) seems to be lower-level-sensory, since the illusion is relatively immune to at least certain cognitive factors, such as a feedback and reward (Andersen, Tiippana, & Sams, 2004; Mishra, Martinez, Sejnowski, & Hillyard, 2007; Rosenthal, Shimojo, & Shams, 2009; Shams, Kamitani, & Shimojo, 2000). This illusion demonstrates that the modality carrying the more discontinuous (and therefore salient)

signal becomes the influential or modulating modality (Shams, Kamitani, & Shimojo, 2002; Shimojo & Shams, 2001). It has also been shown that a wide variety of crossmodal information is combined such that the resulting variance is minimized, thereby mimicking maximum likelihood estimation (MLE) models (Ernst & Banks, 2002). Ernst and Banks were able to conclude from MLE that the modality that dominates in crossmodal information integration is the one with the lowest variance.

Crossmodal intrinsic mappings also bridge modalities with learned spatial, linguistic, or emotional connections (Spence, 2011). Many visual and auditory mappings have been studied for over 50 years, including relating high visuo-spatial position with high pitch and vice versa (Pratt, 1930). Other audio-visual mappings include the matching of loudness with brightness (Stevens & Marks, 1965) and the matching of visual spatial frequency and auditory amplitude modulation rate (Guzman-Martinez, Ortega, Grabowecky, Mossbridge, & Suzuki, 2012). Mappings also exist between vision and tactile sensation (Spence, 2011), including between darkness and weight (Walker, Francis, & Walker, 2010), and color and temperature (Morgan, Goodson, & Jones, 1975). These are likely prior to any associative learning via experiences, considering that these correspondences are common across various natural languages (from different origins) with very few exceptions. In fact, preverbal infants of only 3 to 4 months have been shown to have correspondences between high auditory pitch and high spatial position as well as visual “sharpness” (Walker et al., 2010). Nevertheless, new relations between senses can be learned with training (Ernst, 2007).

Modalities are also plastic after development and can generate learned (or trained) relations across senses, as witnessed in visual activation during echolocation, sound

localization, and braille reading in the blind (late blind *vs.* early blind) (Bavelier & Neville, 2002; Cohen et al., 1997; Collignon, Voss, Lassonde, & Lepore, 2009; Sadato et al., 1996). Braille reading activated primary visual cortex (BA 17) and extrastriate cortices bilaterally in blind participants (Sadato, et al., 1996). Repetitive Transcranial Magnetic Stimulation (rTMS) was used to deactivate visual cortical regions in blind braille experts, and generated errors in braille interpretation (Cohen, et al., 1997). In fact, an early-blind individual lost the ability to read braille after experiencing a stroke in the occipital cortex (Hamilton, Keenan, Catala, & Pascual-Leone, 2000). These results demonstrate a functional and causal link between visual activation and the ability to read braille in the blind. Other studies provide even more evidence for plasticity in the handicapped, such as enhanced visual ERPs (Event Related Potentials) in early-onset deaf (Neville & Lawson, 1987; Neville, Schmidt, & Kutas, 1983), auditory ERPs in the posterior (occipital) region in early and late blind (Kujala et al., 1995), and posterior DC potentials in blind by tactile reading (Uhl, Franzen, Lindinger, Lang, & Deecke, 1991).

Further, due to plasticity-induced changes, it has been proposed that the brain, including the visual cortex, may be “metamodal,” such that brain regions are segregated by processing of different types of information, and not by stimulus modality (Pascual-Leone & Hamilton, 2001). The metamodal theory of the brain was supported by the activation of the shape-decoding region, Lateral Occipital tactile-visual area (LOtv), by audition when shape was conveyed by vOICE encoded sounds (Amedi, et al., 2007).

Crossmodal perceptual organization usually refers to Gestalt principles, such as proximity-based (both in space and time) grouping/segregation, regularity, and Prägnanz (good shape). Vision, audition, and somatosensation have partly the same, but partly

different (unique) perceptual organization rules. For example, segregation or chunking rules operate across modalities in the same way at the most abstract level, but indeed, it could be spatial in vision but temporal in audition (Bregman & Campbell, 1971; Neri & Levi, 2007; Vroomen & de Gelder, 2000). SS provides an opportunity to investigate what would happen to such perceptual organization rules when between-modality connectivity is enhanced by training. To be more specific, questions including: A) would the auditory or the tactile modality acquire vision-like perceptual organization rules and B) would crossmodal combinations themselves self-organize and generate new crossmodal organization principles, can be investigated in detail with sensory substitution.

Existing literature on crossmodal interactions is a guide to understanding and interpreting the visual nature of sensory substitution processing. Sensory substitution also requires plastically generating new learned relationships across modalities, but it may also rely on existing developmental connections. In fact, SS might modulate the strength of existing developmental connections, and thereby alter crossmodal perception, even in sighted participants. Ideally, the training of participants can exploit these existing crossmodal interactions and mappings to enable effortless training and signal interpretation. In addition, training on SS devices should take into account crossmodal interaction variance across both functional and experimental participant groups, including the early blind with no visual experience, the late blind who have limited visual experience, and the sighted with normal visual perception (Bavelier & Neville, 2002; Poirier, De Volder, & Scheiber, 2007).

Perceptual Constancies and Sensory Substitution

Constancies (the perception of a unchanging environment despite various observation conditions such as head movement, lighting changes, and distance alteration) are critical to the perception of stimuli as distal (within the external environment) rather than proximal (on the retinal surface) (Palmer, 1999). This distal nature of perception conferred by constancies is equivalent to the externalization of perception. It is critical to perceive the world distally, as it is useful for the functional engagement of the objects in the environment. For example, if a person would like to pick up a dropped coin, it is easier to recognize the coin if its color and brightness is constant whether it is under the shadowy desk or brightly illuminated in your hand. Further, it is easier to pick up the coin if the size of the coin appears to be constant independent of whether it's closer (such as in your hand) or farther away (such as under the desk). There are several types of visual constancies such as size constancy (constant size independent of distance), shape constancy (constant shape independent of object rotation), orientation constancy (constant object orientation independent of head tilt), position constancy (constant object position independent of head and body movement), and brightness constancy (constant brightness independent of external illumination) (Palmer, 1999). Auditory perceptual constancies exist as well, such as: Loudness constancy (constant loudness independent of distance, like size constancy) (Zahorik & Wightman, 2001) and rotation constancy (constant sound location independent of head rotation, similar to position constancy) (Day, 1968).

Two general ideas dominate the visual perceptual constancy theory: the indirect idea originated by Helmholtz, and the direct idea generated by Gibson. The indirect

approach claims that constancy is constructed by using the retinal image information in combination with other additional information (optical and nonoptical), such as accommodation in size constancy (von Helmholtz, 1925). Additional information for constancy processing can include head angle for orientation constancy, or external lighting properties for brightness constancy. In fact, for position constancy, it has been argued that an efferent copy of movement commands is sent to brain regions for constancy computation. In contrast, the direct approach uses image properties that are invariant to determine constancy (Palmer, 1999). For example, position constancy can be generated using optic flow calculations based on the retinal image. In particular, Gibson argues that some constancy properties such as size and distance in size constancy are unavoidably linked and inseparable (Gibson, 1950a).

While people can perceive object constancies, they can also perceive object changes. These two types of perception (constancies and changes) are known as the distal and proximal modes (respectively) (Palmer, 1999). The distal mode (*i.e.*, using constancy) is useful for functional tasks and interaction with environmental objects (as discussed above). Therefore, the distal mode frequently dominates in daily life tasks. The proximal mode, or retinal image perception, can be occasionally useful for select activities such as realistic painting or the detection of an illusion due to distal (or constancy) assumptions. Occasionally, individuals use the proximal mode of vision for these tasks, and can even perceive constancies and changes simultaneously.

Constancies processing has yet to be tested with vOICE or with sensory substitution devices. If it can be proven that participants can learn constancy with sensory substitution, it will be an important indication of perceptual externalization and

distal perception with sensory substitution devices. Further, constancy will be important to the comparison of sensory substitution processing to purely visual and purely auditory processing capabilities.

Visual Attention, Automaticity, and Sensory Substitution

Vision is well known to be automatic, requiring limited attention and cognitive processing for complex tasks such as visual search (A. Treisman, 1985) or the processing of faces (Palermo & Rhodes, 2007). Overt visual attention is often expressed with eye and head orienting movements that bring the object of interest into view for inspection, such as with facial feature evaluation. However, attention can also be exerted by internal mechanisms of focus that highlight different object properties or spatial regions, called property selection and spatial selection, respectively (Palmer, 1999). In addition, attention can be oriented based on exogenous cues (cues from the environment, such as a loud sound) or on endogenous cues (internally generated cues) (Posner, 1980).

Attention selection and processing without attention (automaticity) has been studied in audition and vision with distraction paradigms. The goal of distraction is to focus attention on an unnecessary task, thereby limiting the attention that can be used in the task of interest. In audition, distraction tasks were performed using repetition of words communicated through the left or right side of a pair of headphones (Cherry, 1953; Palmer, 1999). Then participants were tested for the information heard in the unattended headphone. Participants were able to detect general sensory information in the unattended ear, such as the voice gender or if it was speech, but not specifics such as the language or content. Further experiments showed that personal information could also be detected in the unattended ear, such as the individual's name. The combined early and

late processing required for general sensory information detection as well as detailed personal information detection generated the attenuator theory of auditory processing. This theory claims that attention is allocated for early processing information of interest as well as highly salient representations of words (Palmer, 1999; A. M. Treisman, 1960). The threshold for conscious awareness of words processed is then dynamically regulated based on an individual's value of that word. Visual distraction paradigms were designed with a similar distraction task (such as relative length of two presented lines) and with an unattended visual stimulus of interest. When questioned about the unattended object, participants could identify the location, color, and number of objects, but not the shape (Palmer, 1999; Rock, Linnett, Grant, & Mack, 1992). However presentation of the object's name did allow for most participants to identify it although it was unattended (Palmer, 1999). Therefore, similarly to audition, a combination of early and late processing is allocating attention in vision.

Several automaticity-imaging studies have indicated that visual stimuli are still processed in the brain despite being cognitively ignored, via a distraction paradigm. Schwartz and colleagues showed that the attention load of the central task modulated the visual activation of the ignored peripheral stimuli (Schwartz et al., 2005). Further, Schwartz *et al.* indicated that peripheral regions neighboring the central visual regions (processing the attentive task) were suppressed more by high attention load than distant peripheral cortex. Several neuroscientists have investigated emotional processing of face stimuli with limited attention, and have obtained mixed results (Pessoa, 2005; Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Vuilleumier, Armony, Driver, & Dolan, 2001). Pessoa claims that these mixed results are in fact logical, if cognitive load is

considered, with the high cognitive load reducing the emotional vision processing the most, and vice versa (Pessoa, 2005). However, in the words of R. Palermo, “although neural responses in face-selective cortex are reduced for ignored compared with attended faces, they are certainly not eliminated” (Palermo & Rhodes, 2007). Results seem to indicate that unattended visual stimuli have reduced activation in visual cortex relative to attended stimuli, but that nevertheless unattended stimuli are still processed in visual regions.

While vision automaticity (processing without active attention) has its limitations, it does occur for simple image/sound properties, and complicated properties of personal value. These visual automaticity properties are important to the hierarchy of visual processing and will be important for the comparison of vOICe to visual processing. Unfortunately, the current sensory substitution imaging studies all use tasks with active attention and cognitive processing. No distraction task paradigms with SS have been tested with psychophysics or brain imaging. Therefore, the intensity of attention required and the different roles attention may play in interpreting vOICe, or other sensory substitution devices remains unknown. While it has been assumed that vOICe is processed entirely cognitively with active attention resources, this may not entirely be the case. Elements of vOICe processing, mediated by existing crossmodal interactions and connections, may be processed without attention in a similar manner to that in which some elements of vision are processed. Automaticity and its relevance to sensory substitution will be discussed further in Chapters 3 and 4.

CHAPTER 2:

LENGTH AND ORIENTATION CONSTANCY LEARNING

WITH AUDITORY SENSORY SUBSTITUTION

Introduction

Chapter 1 discussed that depth illusions (Renier, Laloyaux, et al., 2005), as well as natural and artificial object identification and localization (Amedi, et al., 2007; Bachy-Rita, et al., 1998; Poirier, De Volder, Tranduy, et al., 2007; Proulx, et al., 2008) including face and word identification (Plaza, et al., 2009; Striem-Amit, Cohen, Dehaene, & Amedi, 2012) have been shown to be learned by SS users. Whether it only indicates sensory-motor learning, or rather adaptive changes of some intrinsically-visual quality/function, remains unsolved. A basic problem with SS is that perceptual constancy, the ability to perceive a feature as constant despite changes in a dynamic visual scene, has not been investigated or shown to be learned with SS. As detailed in Chapter 1, constancies are a critical element of perception that are important to functional task performance as well as accurate environmental perception (*i.e.*, externalization). If constancies can be learned, it would indicate the potential for SS users to attain a high level of functional capability as well as SS to behave as a perceptual modality (like vision or audition alone). Therefore, we used the vOICE auditory SS device to demonstrate that the sighted and blind can learn orientation and length constancy tasks, and that this learning is amplified by dynamic interaction with stimuli, providing further insight into how visual-motor experiences shape perceptual constancies in general.

Visual orientation constancy is the ability to estimate the angle of an object independent of head tilt. This constancy is useful for determining an object's angle relative to gravitation (Palmer, 1999). It would be particularly useful to the blind SS users, as it would allow them to determine object stability via the object's tilt, thus perceiving it as an external object independent of their own locomotion/movement. The stability of tables, chairs, and other furniture as well as natural objects such as rocks, trees, and branches is important mobility information for the blind. It is also useful for obstacle avoidance of leaning objects whose position in space (determined via orientation constancy) is critical to locomotion around them. In particular, orientation constancy allows detection of low-hanging branches that often hang at an angle, and are undetectable by a cane grazing the ground. If a branch angle were misinterpreted by the lack of orientation constancy, than the blind user of a SS device would collide with that branch. Therefore, orientation constancy is particularly relevant to SS users, and is valuable to daily functioning with SS.

Visual orientation constancy is generated by proprioceptive information of head orientation, kinesthetic feedback, and visual frame of reference, allowing for the correction of tilted images due to head tilting (Palmer, 1999). Object orientation is first identified by V1 orientation sensitive cells that detect the angle of lines as they appear on the retina. However, if these cells alone determined perceived orientation, then objects would rotate with head motion. To remove head motion from the perceived angle of objects, vestibular organs sense head angle (via utricle and saccule organs) and changes in head angle (via semicircular canals). Further, kinesthetic feedback provides information about body movement (from sensors in joints) that is used in orientation

constancy. Finally, visual frame of reference or context informs orientation constancy. Frame of reference is a set of heuristics of the typical vertical angle of particular objects (such as walls forming a room). Frame-of-reference violations can even generate illusions (such as a tilted room illusion) that override proprioceptive information (Shimojo, 2008). In particular, a rod and frame illusion causes a central line or rod to be perceptually tilted by the context of an external tilted rectangle or frame. Due to frame of reference, in a lit environment orientation constancy is quite robust; however, in a dark room, an approximate error of 10 percent can occur at large head tilts due to the loss of visual context (Palmer, 1999). Overall, both non-visual information and environmental visual context contribute to the accurate perception of orientation constancy.

The cortical processing of orientation constancy has been studied with a variety of imaging techniques and lesion patients. Corbett and colleagues used event-related potentials (ERPs) to measure the neural processing underlying the rod and frame illusion (Corbett, Enns, & Handy, 2009). They found that neural processing represented by P3, characterizing later processing, mediated orientation constancy. A lesion case study indicated a “room tilt illusion” for a patient with posterior cortical atrophy. The scientists hypothesized that the illusion was due to disordered processing of vestibular and visual inputs (Crutch et al., 2011). Finally, Denny and Adorjant showed that electrophysiological responses of cat primary visual cortex were modified by head rotation (Denney & Adorjanti, 1972). Overall, these results indicate later visual processing of orientation constancy as well as possible feedback modulation of earlier visual regions for constancy (in cats).

Length constancy (the ability to estimate length independent of object angle) is a sub-type of shape constancy (the ability to estimate object shape independent of perspective). Length constancy is particularly difficult to accomplish with the vOICe SS device, due to the vOICe image-to-sound encoding of sound frequency in the vertical dimension and scan-time laterally (Meijer, 1992). Horizontal line length is encoded by duration of the sound, whereas the vertical line length is encoded by the range of frequencies in a very brief sound. Inevitably, lines of different angles but the same length are not only perceptually quite different with vOICe, but the computation of length estimation is different as well. In vision, the retina can estimate line length with the same neural computation in any angular dimension. Therefore, it will be particularly interesting if participants can overcome these challenges to learn length constancy with the vOICe device.

Shape and length constancy have been investigated extensively in visual perception. Two-dimensional shapes rotated in three dimensions have been shown to have robust visual shape constancy in both adults and infants (Palmer, 1999). This result indicates that shape constancy is surprisingly innate but also dependent on adequate depth perception. Length constancy of 2D objects rotated in the 2D plane (used in this experiment) is a simpler case, which largely avoids this depth perception dependence. In general, shape constancy is constrained by a tendency toward symmetrical shape perception, which aids in shape constancy of symmetrical shapes and limits constancy of non-symmetrical shapes. Shape constancy of 3D wire objects and unfamiliar opaque 3D objects is surprisingly limited, and is an example of proximal perception (discussed in Chapter 1). Despite this, shape constancy of 3D familiar objects seems to be easy for

sighted people. For example, recognition of a banana from 4 different perspectives is straightforward for the typical sighted individual. The ease of daily-life shape constancy could derive from familiarity with the object, or from gradual shape changes as an object moves, or even from identifying axes of symmetry. Overall shape constancy of simple 2D shapes is trivial for the sighted, whereas constancy of 3D arbitrary shapes is difficult and problematic.

fMRI imaging studies of shape constancy have indicated that it is processed in several visual cortical brain regions. Vuilleumier and colleagues used a neural fatigue paradigm with fMRI imaging to show that the left fusiform region decreased in activity in response to repeated stimuli of varying stimuli viewpoints and sizes (Vuilleumier, Henson, Driver, & Dolan, 2002). This fatigue paradigm indicates that the left fusiform is processing visual information from multiple perspectives (*i.e.*, view invariant), a required element of shape constancy. In contrast, the right fusiform activity was only fatigued in response to repeated stimuli from a single vantage point. Kourtzi *et al.* also studied 3D shape perception with a fatigue fMRI design, but their experiment was focused on the lateral occipital complex (LOC), a region known to process shapes (Kourtzi, Erb, Grodd, & Bulthoff, 2003). It was determined that the LOC fatigued in response to objects with the same 3D shape but different 2D shapes. The LOC did not fatigue in response to objects with the same 2D shapes but different 3D shapes. In other words, the LOC was coding for 3D shape in real world coordinates rather than the 2D retinal image shape. Finally, an fMRI study contrasted identity and orientation tasks for objects oriented in depth. Their results “suggested that the parietal/frontal object areas encode view-dependent visual features and underlie object orientation perception” (Niimi, Saneyoshi,

Abe, Kaminaga, & Yokosawa, 2011). This object orientation perception is likely an input to the shape constancy that must integrate all views of an object to be view-invariant (performed in a region such as left fusiform region, as detailed above). The neural processing of shape constancy has been shown to involve parietal/frontal areas and the right fusiform region in view-dependent perception that is then processed by the LOC or the left fusiform region for view invariance.

As discussed in Chapter 1, constancies (including orientation and length constancy) may be considered a basis for object externalization or distal perception. Orientation constancy allows the object angle to be perceived not as it is on the retina, but instead as an angle of an independent, *external* object in real-world coordinates, thereby enabling adaptive behavior. Learning orientation constancy in SS will therefore be critical to externalizing this new type of sensory input. Length and shape constancy allow an object to be recognized not as changing identity after rotation, but rather as a cohesive single object in the environment. This allows for the object to be externalized out in space as a real singular object. Stimuli externalization is critical for adaptive behavior, because objects are perceived as they are positioned, oriented, and shaped in the external environment rather than on the retinal image. For example, externalization of stimuli will allow for a blind individual to correctly locate a drop cane, approach the cane, and pick up the cane. Without orientation and position constancy, the cane would be jittering in space as the individual moved and would be impossible to reach and grasp. Without shape constancy, the cane would appear to be a different object when the participant tilted their head, making recognition of the cane quite difficult. Overall constancies are critical to the functional use of SS and to improving SS device usage.

Methods

Twelve blindfolded sighted and four blind participants (three late blind, one congenital) were trained on the vOICe device for at least 8 days at approximately 1 hour per day performing three evaluation tasks.

To evaluate orientation constancy, participants were presented with a bar at 6 different angles (0, 90, 45, -45, 22 or -22 degrees relative to vertical) with three potential head positions (vertical, tilted left, or tilted right) and determined the angle of the bar. The experimenter placed the bar on a black felt-covered wall in front of the seated participant, and visually estimated each angle position to be presented to the participant. Participants were permitted to determine the head tilt that they were most comfortable using in each trial, provided their head was stationary. One head position was requested for each trial.

To evaluate the precursor to length constancy, participants were presented with 5 lengths of bars (5AFC: 9, 12, 15, 18, 21 cm), while the bar was placed in one of four orientations (0, 90, 45, or -45 degrees relative to vertical). Participants were asked to determine the length of the bar presented independent of the angle it was presented at. Since our primary aim was to explore training style/design, participant training was varied to determine the optimal training procedures. Two sighted participants were directed not to use head tilt during the length constancy task, and were not included in Figure 2.01 or Figure 2.05. One sighted participant was directed to use head tilt intermittently (4 out of 12 sessions with head tilt), and was excluded from Figure 2.05, but included in Figure 2.01. The remaining participants were asked to tilt their head in

the initial trials and at the end of the training. Two participants evaluated the bar length without head tilt. Figure 2.05 includes and excludes different participants in a few data points so that the figure can show data for head tilt trials in sessions 0-5, and no-head-tilt data in trials in session 6-7 (sessions 6 data point = 5 participants, session 7 = 7 participants, all other data points = 9 participants). Figure 2.05 was designed with the head-tilt in initial sessions and no head tilt in later sessions to show the retention of the learning gained during the head tilt sessions, which is important to our argument that head-tilting aids learning. The same procedure was used for Figure 2.06 so that it can show data for head tilt in sessions 0-7, and only data for no-head-tilt in session 8-9 (session 8 data point = 3 participants, all other data points = 4 participants). Due to the technical constraints in the usage of the device, the desire for training exploration, and various practical and cognitive limitations in blinds, we could not carry out the experiments just unanimously with one simple and identical procedure.

If time permitted, the participants performed some other tasks such as object recognition and localization, whose data we decided not to include here. Both experimenters followed the same training protocol (training procedures detailed day to day) outlined for all experiment tasks (details in Appendix B).

Participants used a vOICe device to learn the constancy tasks. The vOICe device used a camera embedded in a pair of sunglasses or a webcam attached externally to glasses. Sighted participants were requested to close eyes during training and evaluation and wore opaque glasses and/or mask. The camera provided live video feed of the environment, and used a small portable computer to encode the video into sound in real time. The vOICe device translates the horizontal spatial dimension to scan-time, the

vertical spatial dimension to frequency range, and image brightness to sound loudness. The vOICe software was obtained online at seeingwithsound.com and used for the video to sound encoding. All training sessions were recorded for later data analysis and/or presentations; participants were notified of video recording.

Three different vOICe device camera setups were used during training. All setups had a camera attached or embedded in glasses, a small portable computer connected to the camera, and earphones (either separate from the glasses or attached). The camera could be on the side of the glasses, or in the center on the bridge of the nose. Sighted participants' natural vision was obscured with black felt covering the glasses, or a sleeping mask worn under the glasses. In principle, these technical differences would not make any difference in terms of training efficiency and task performance, except a possible minor difference in spatial perception due to the gain of the camera, camera field of view and camera placement. Two blind participants were forced to transition from one camera and device setup to another setup partway through training due to device failure; their data did not indicate any difficulty with this transition.

Data analysis of head tilt and time to decision were performed on the video recordings of training sessions. Head tilt was quantified by counting the number of trials in which the participant used head tilt while exploring the stimulus, and then divided by the total number of trials (160 trials for sighted and 200 trials for blind). Head-tilt was estimated for training sessions with missing video by using the average head-tilt for sessions of the same type (such as head tilt allowed or head tilt not allowed). Two participants were trained without any head tilt permitted in the training, and they are included in the head-tilt correlation and time-to-decision plot but not the main data set on

performance, to prevent inhomogeneity in training.

Data analysis on time-to-decision was performed by recording the onset and end of a task during the training session for all training sessions of all training participants. The data was averaged across sighted participants and across blind participants. Training sessions lacking a video recording were omitted from the analysis. One blind participant on the orientation constancy task was omitted from the time-to-decision data due to the lack of three consecutive training session videos; no other participants lacked three or more consecutive training session videos.

ANOCOVA and regression analyses were performed in MATLAB using the `aoctool`, `regstats`, and `glmfit` functions.

Results

Sighted and blind vOICe users were able to classify line angle independent of head tilt (orientation constancy), and to learn to further improve (Figure 2.01 and Figure 2.02). The rate of improvement was significant in both groups (Sighted, 8 training sessions: $p < 0.00$; Blind, 9 training sessions: $p < 0.00$). Blind participants had an average slope of improvement that was not significantly different from that of the sighted ($p_{slope} < 0.195$). However, the intercept of the improvement curves was significantly different between sighted and blind users ($p_{intercept} < 6.54 \times 10^{-9}$) with the sighted starting training at a higher percent correct, likely due to the blind users' diminished spatial perception.

Orientation constancy performance was also evaluated separately at each head position (vertical, tilted left or tilted right) (Figure 2.03 and 2.04). Head vertical position

outperformed the head-tilted-left and head-tilted-right conditions for both the sighted and blind participants. The head-vertical position had the advantage of no angle shift calculation; in other words, the angle heard by the participant via vOICE is the angle in the environment, which was not true for the head-tilted-left or right conditions (also, the majority of training experiences had been at this angle). Therefore, with the angle directly perceived and no arithmetic added to the task, the head-vertical task was easier to perform than the head-tilted-left or right task. ANOCOVA analysis indicated that no slope pair between the vertical-head slope, head-tilted-left slope and head-tilted-right slope was significantly different in the blind or sighted participants. Intercepts were also evaluated with ANOCOVA analysis (assuming the slopes are equivalent), and the head-vertical condition was significantly different from the head-tilted-left ($p_{intercept} < 0.0002$) and the head-tilted-right conditions ($p_{intercept} < 1.62 \times 10^{-6}$) for the sighted participants. The blind participants had significantly different intercepts for vertical compared to head-tilted-left ($p_{intercept} < 0.0008$) and vertical compared to head-tilted-right ($p_{intercept} < 0.0036$).

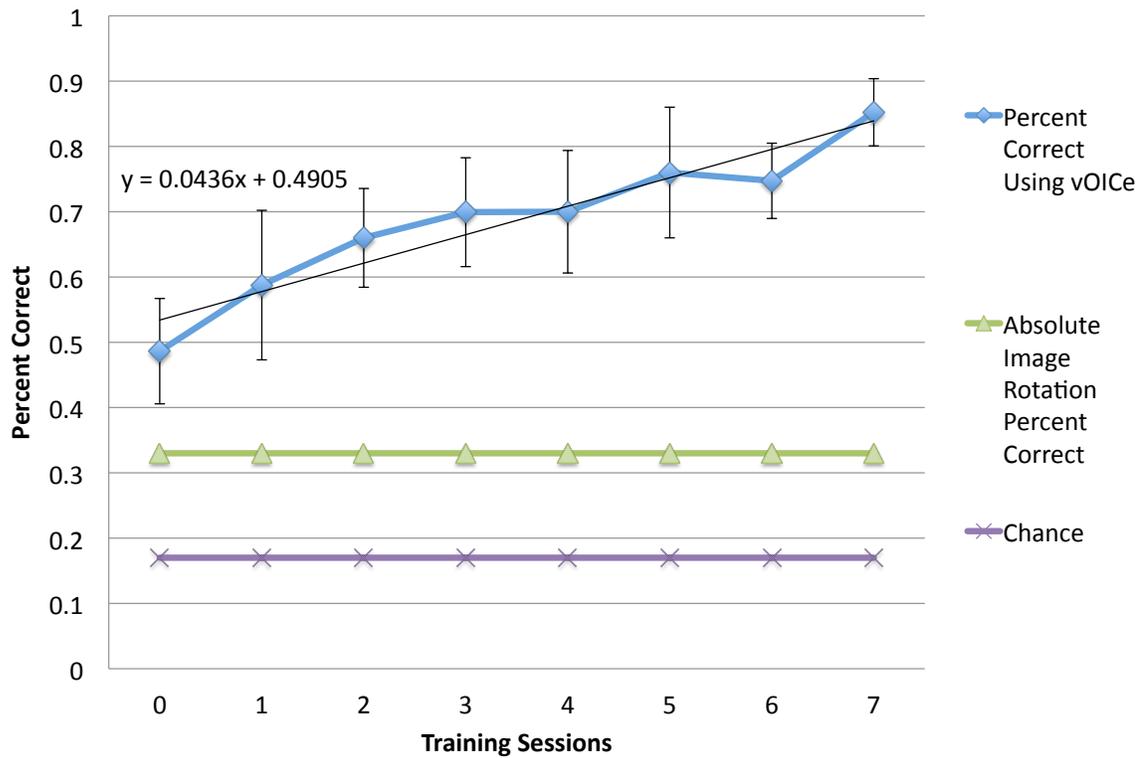


Figure 2.01. Sighted orientation task data. Performance in the orientation constancy task (classification of line angle independent of head tilt) as a function of the number of training sessions in the sighted participants ($N = 10$). Error bars are the standard deviation. Blind participant data are in a Figure 2.02. (Note: Two sighted participants excluded due to differences in training).

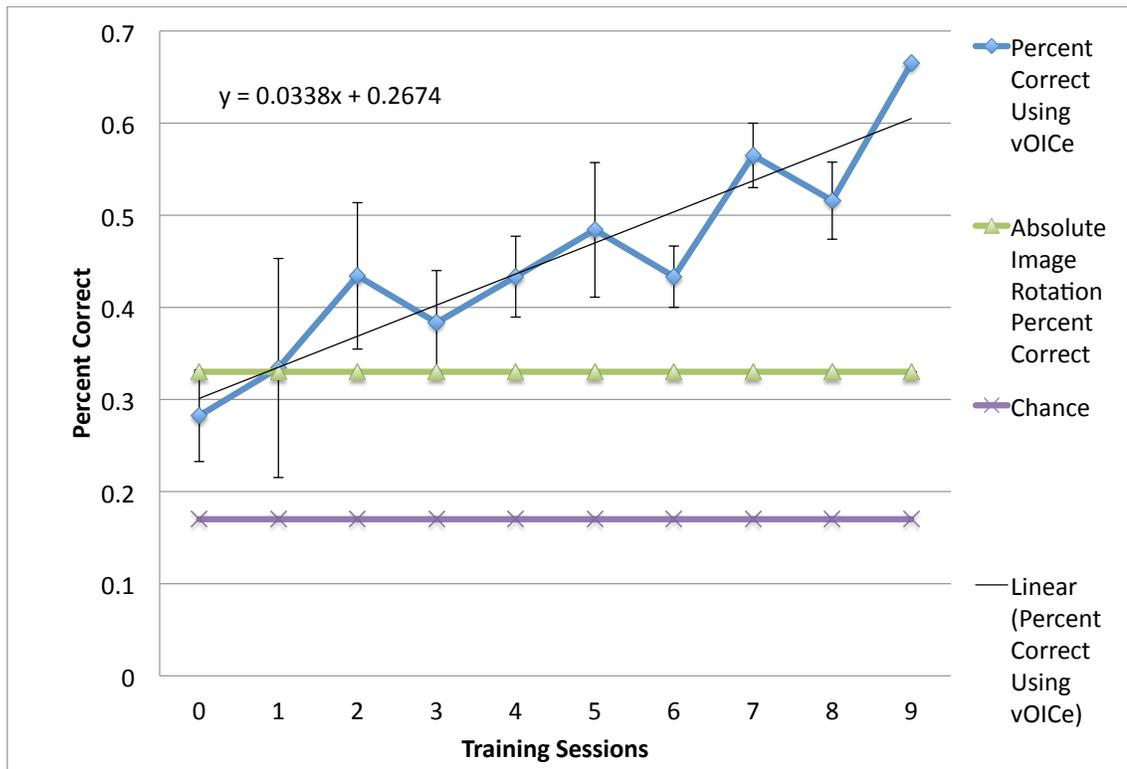


Figure 2.02. Blind orientation task data. Task performance of orientation constancy as a function of the number of training sessions in the blind participants ($N=4$). Error bars are the standard deviation. The Absolute Image Rotation Percent Correct in Figure 1A is the percent correct if the angle of head tilt is unknown (*i.e.*, only head vertical can be correctly identified, or 1/3 correct).

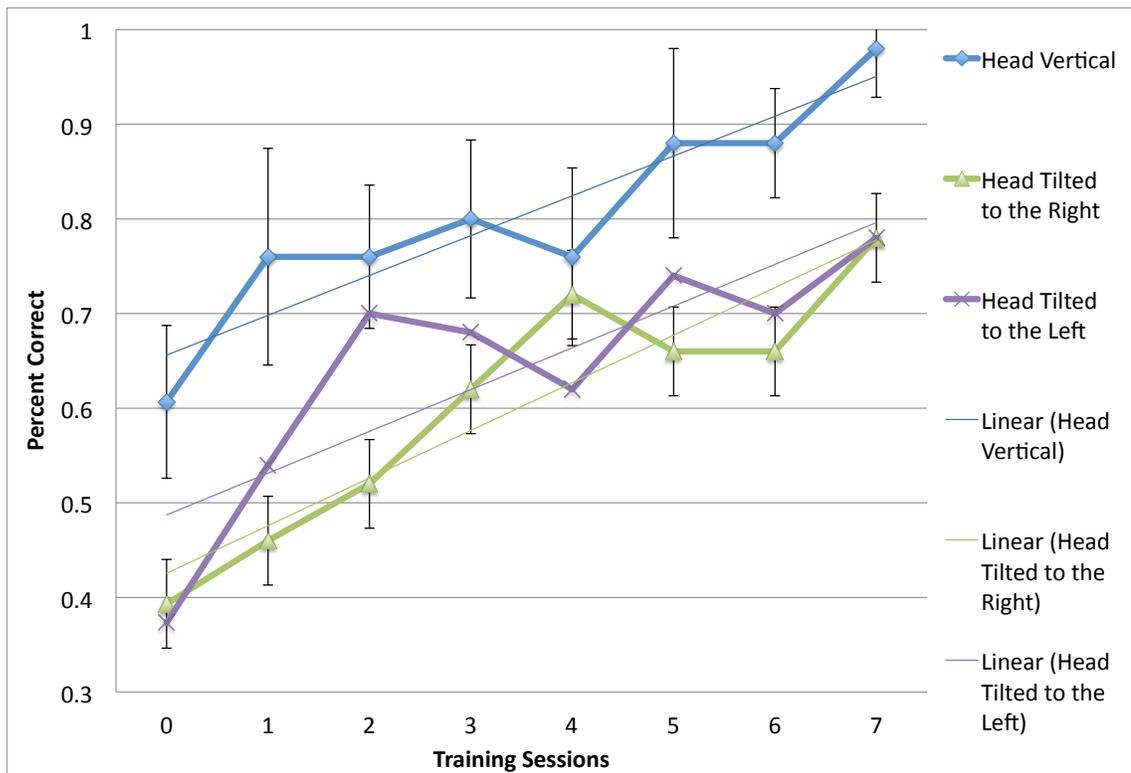


Figure 2.03. Sighted orientation task data divided by head tilt. Task performance of orientation constancy as a function of the number of training sessions in the sighted participants ($N = 10$). Data is separated into the participants' percent correct for each of the potential head positions: Vertical, tilted left, or tilted right. Error bars are the standard deviation.

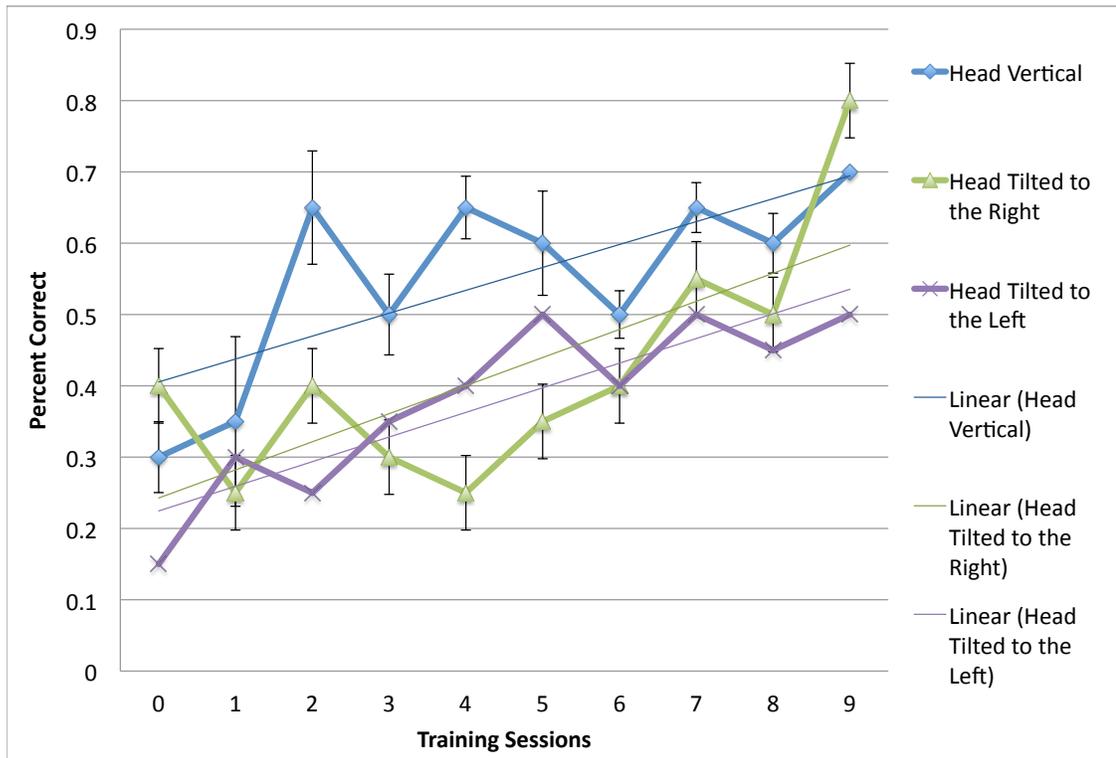


Figure 2.04. Blind orientation task data divided by head tilt. Task performance of orientation constancy as a function of the number of training sessions in the blind participants ($N=4$). Data is separated into the participants' percent correct for each of the potential head positions: Vertical, tilted left, or tilted right. Error bars are the standard deviation.

Sighted and blind vOICe users were able to classify line-length independent of angle (length constancy), and to learn to further improve (Figure 2.05, Figure 2.06). The rate of improvement was significant in both groups (Sighted, 8 training sessions: $p < 0.01$ Blind, 10 training sessions: $p < 0.03$). Nonetheless, blind participants had an average slope that was not significantly different from the sighted, while the intercepts were significantly different ($p_{slope} < 0.179$, $p_{intercept} < 0.0014$). During head-tilt allowed sessions, head-tilting was encouraged. Head tilting frequency during the task correlated significantly with improved line length classification (Figure 2.07). The sighted participants correlated head tilt with length constancy task improvement with a coefficient of 0.6560 ($p < 0.03$), whereas sighted and blind participant data combined had a coefficient of 0.6024 ($p < 0.02$). The blind-only correlation is seemingly lower partly because the participants were fewer, and they have a wider range of capabilities and spatial perception. (For further comparison of blind to sighted, see Figure 2.08). Head tilt frequency was not significantly correlated with participants' initial performance at the length constancy task (*i.e.*, intercept) (for sighted and blind combined $\rho = 0.0786$, $p < 0.78$).

Blind participants were divided into late and early blind categories, and their slope of improvement was compared to the sighted participants (Figure 2.08). In both the shape constancy and orientation constancy tasks, the early blind participant ($N = 1$) improved the slowest (*i.e.*, smallest slope). The late blind participants ($N = 3$) improved the second slowest, and the sighted ($N = 9-10$) improved the fastest (*i.e.*, largest slope). As discussed in Chapter 1, the sighted participants have the advantage of familiarity with visual principles (such as relative size), and visuomotor skills, due to daily visual

experience. These existing skills can be used to advantage when learning to process the vOICe visually or re-learn a constancy. The late blind have less visual experience than the sighted, as they have been visually deprived for years if not decades. Finally, the early blind have no visual experience. Therefore, the rate of learning seems to correlate with visual experience; however, no definitive statement can be made due to the low number of late blind ($N = 3$) and the extremely low number of early blind participants ($N = 1$).

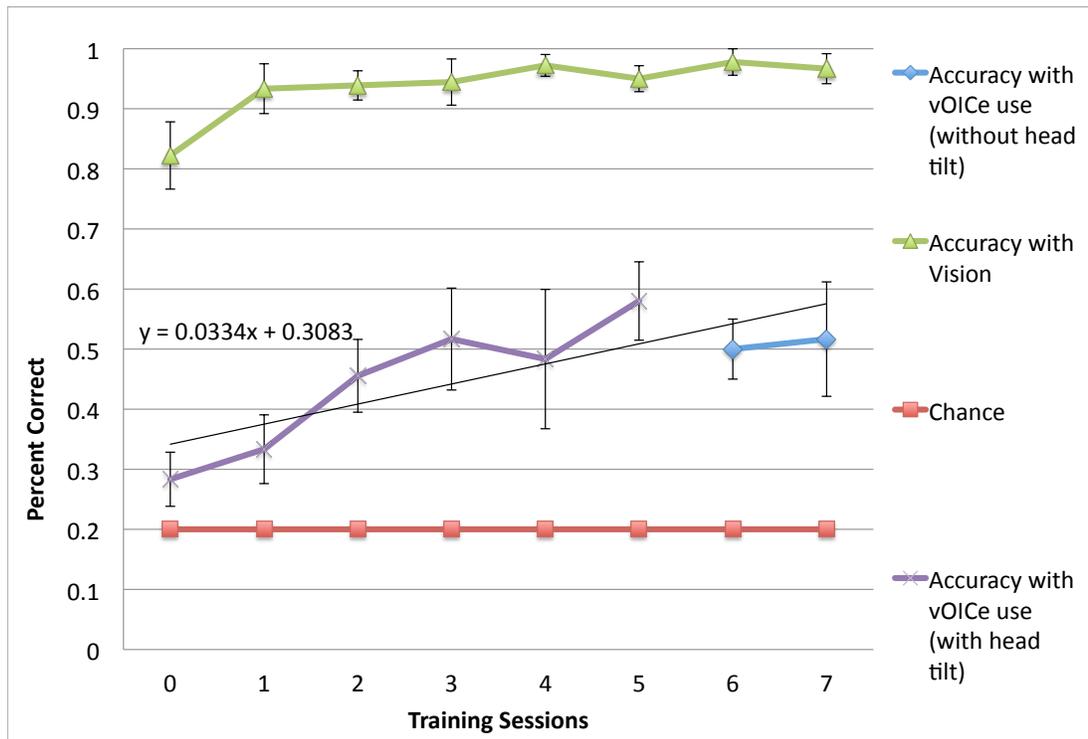


Figure 2.05. Sighted length constancy data. Performance in the length constancy task (classification of line length independent of angle) as a function of the number of training sessions in sighted participants ($N=9$). Error bars are the standard deviation. Blind participant data are in a Figure 2.06 (Note: Three sighted participants were excluded due to no head tilt used, or intermittent use of head tilt as directed by the experimenter).

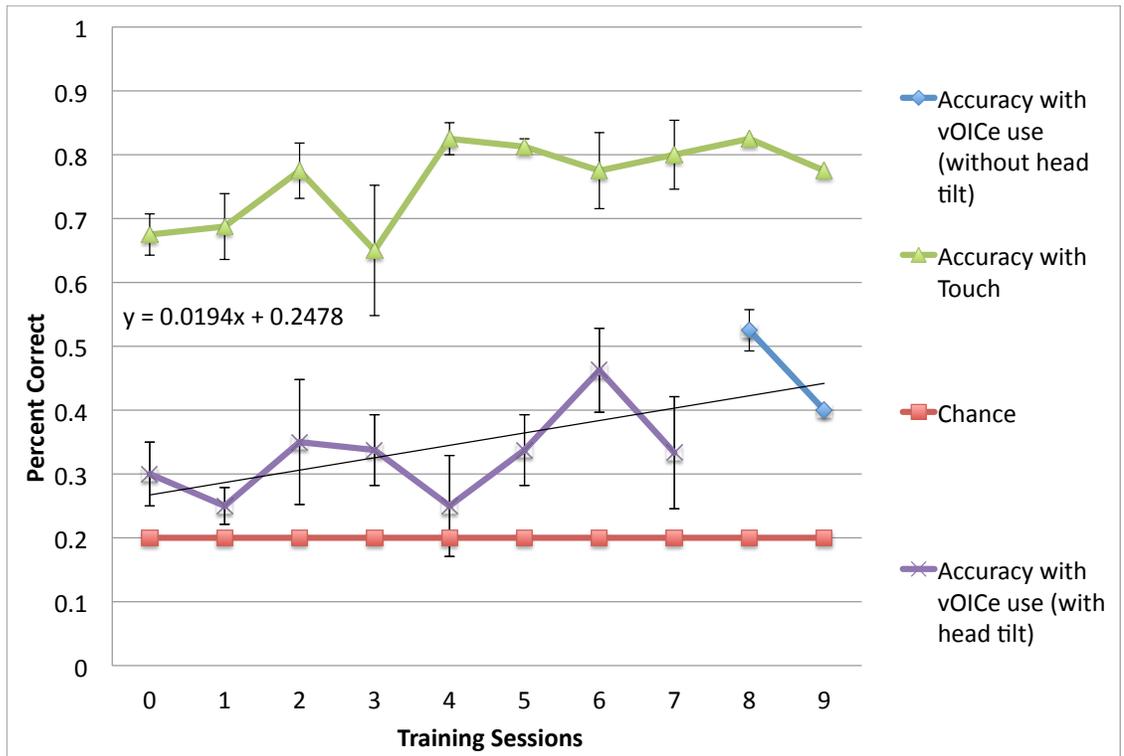


Figure 2.06. Blind length constancy data. Task performance of length constancy as a function of the number of training sessions in the blind participants ($N = 4$). Error bars are the standard deviation.

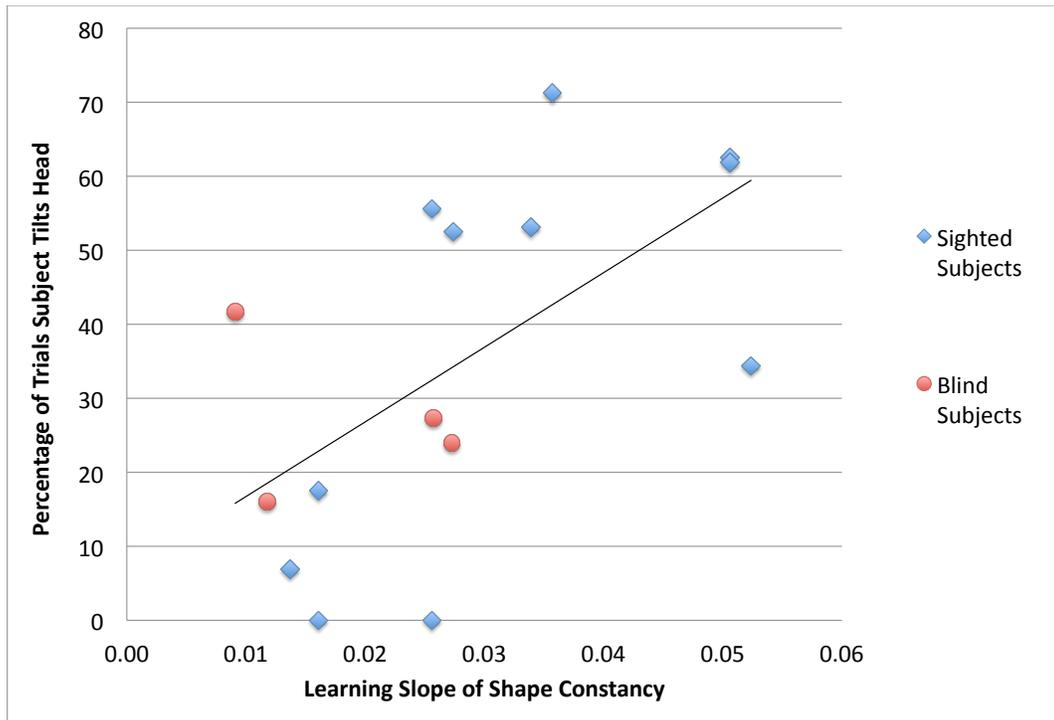


Figure 2.07. Length constancy head tilt and performance improvement correlation. Significant correlation between head tilt and performance improvement in the length constancy task ($N = 16$) ($\rho = 0.6024$, $p < 0.02$). The number of trials that participants tilted their head was counted for all task sessions from video recordings (the average number of head tilt trials were used for sessions with missing video). The percent of trials that head-tilt was used was plotted against the slope of the participants length constancy improvement (from the interpolated slope in individual participant plots similar to Figure 2.05). (Note: One sighted participant was excluded due to intermittent use of head tilt, and two extra sighted participants were included who did not use head tilt at all, and were thus excluded from the analyses for Figure 2.01 and Figure 2.05).

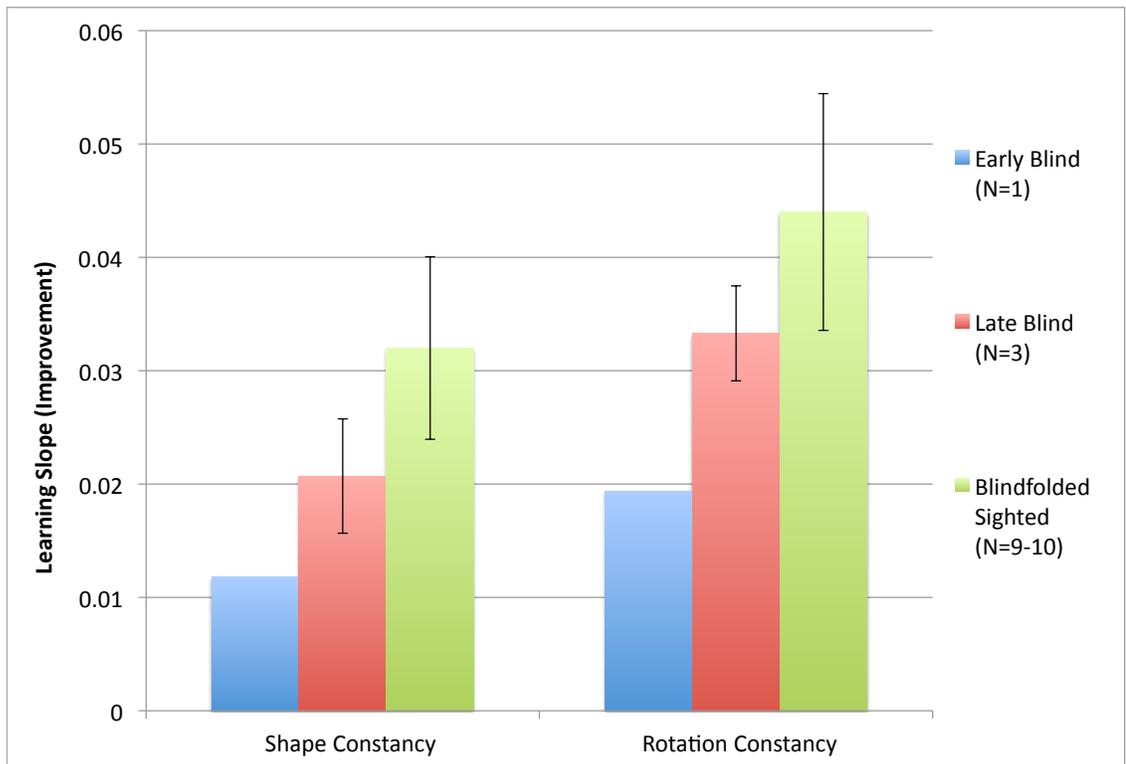


Figure 2.08. Vision and blind task improvement comparison. The slope of task performance as a function training session was determined for each of the blind and sighted participant groups. Slopes were calculated for each individual's data, and then averaged into a group.

The performance time in both tasks decreased as training sessions progressed for sighted and blind participants (Figures 2.09 and 2.10). The decrease in time to perform the training task indicates a tendency toward task automaticity and away from extensive top-down attention, thereby beginning to mimic the intuitive and automatic nature of perceptual constancies in the sighted.

It is revealing that head-tilt significantly correlated with the improved length constancy performance (Figure 2.07). This is the most critical and core finding of this study, as it indicates an improved SS training technique with additional sensorimotor interaction. It also indicates a key method for learning of constancy in vision as well as with SS.

The benefit of head-tilt with length constancy can be described in mathematical and psychophysical terms. As a participant spontaneously tilts their head, they alter the tilt of the camera attached to glasses on the head, thereby altering the angle of the line heard. As the angle of a line is rotated, the length and the width also change according to $L*\sin(\theta)$ and $L*\cos(\theta)$, where L is the length of the line and theta is the tilt, in the head or frame of reference. The change of vertical length (range of pitch in the SS device) and horizontal width (duration of sound) with head tilt are plotted in Figure 2.11 and 2.12, for a line placed vertically (Figure 2.11) or horizontally (Figure 2.12). In these plots, the radius of each half circle represents the line length, which remains the same (*i.e.*, constant) across all the different head tilts or line angles. The brain learns from association of different points on each line to identify each line as one entity (*i.e.*, a white bar of a particular length), and to separate the different lines as separate bars of different length. Learning general length constancy (not just bars used, but all bars at different

angles) requires similar exploration with head-tilt and similar association of different input patterns to be identified as the same real-world object (and it applies to the real world, natural seeing as well as via the sensory substitution device). Further, participants can begin to associate all of the stimuli types (*i.e.*, 0, 90, 45, -45 degrees) for a given line length, which in effect provide orientation-invariance and correspond to the object identity (line length). Obviously, active head-tilting and its sensory feedback have a critical role in such a dynamic associative learning of length constancy, specifically in sensory substitution but more generally in vision. Further, due to the significant head-tilt correlation, memorization has been shown to not be a successful learning strategy for length constancy (*i.e.*, if memorization were used by participants, head tilt would make the task more difficult, whereas head tilt improved performance at the task).

By tilting their head, the observer receives dynamic yet systematic changes of input parameters, as illustrated in Figure 3-A and -B. In effect, learning aims to identify all the data points within each curve as an “identical horizontal line,” whereas discriminating across different curves as “different length.” Quite intuitively, it would be much easier if the brain compared an entire curve *vs.* another in the graph using head-tilt to move along the curve, as opposed to a point-by-point comparison in a set of (static) parameters. One may easily implement this more computationally in terms of S/N ratios in a Bayesian or a MLE framework.

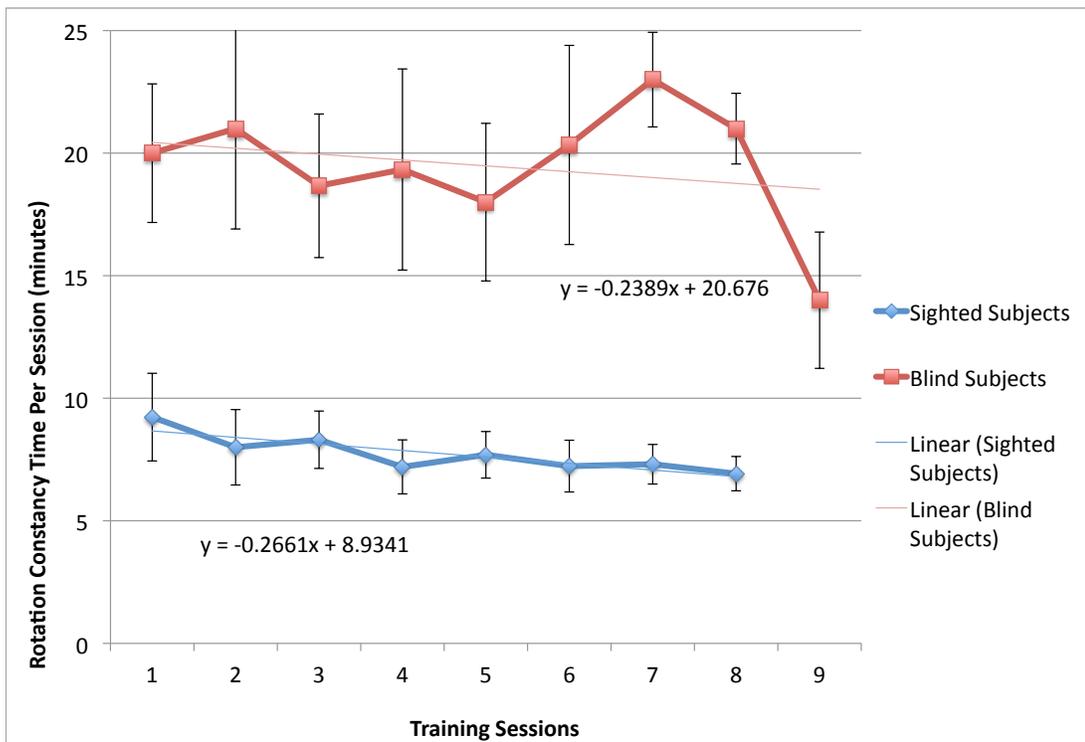


Figure 2.09. Orientation constancy task duration. Orientation constancy task duration for sighted ($N = 12$) and blind participants ($N = 3$). The duration of all trials for each participant was determined from video recordings of training sessions, and then averaged across participants. This orientation constancy task duration is plotted as a function of the training sessions. One blind participant was omitted from the data due to lack of video from three consecutive sessions. Error bars are the standard deviation.

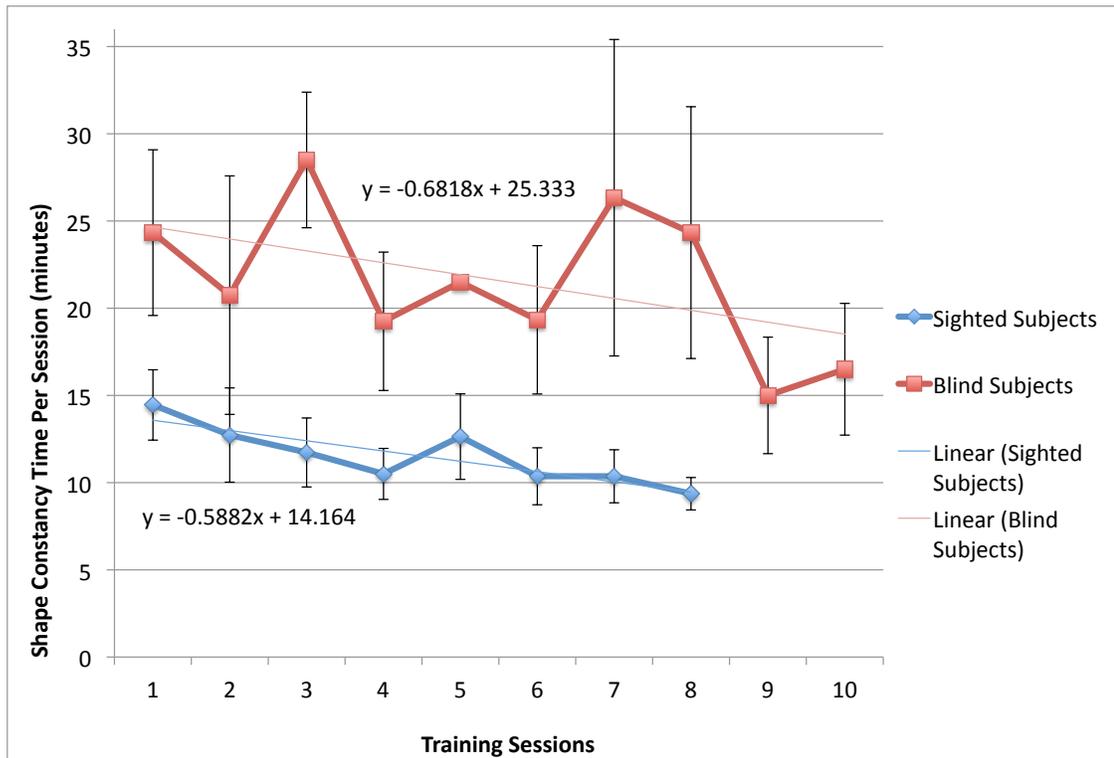


Figure 2.10. Length constancy task duration. Length constancy task duration for sighted ($N = 11$) and blind participants ($N = 4$). The duration of all trials for each participant was determined from video recordings of training sessions, and then averaged across participants. This length constancy task duration is plotted as a function of the training sessions. Error bars are the standard deviation. One sighted participant was omitted due to four different transitions between head-tilt allowed trials and no-head tilt allowed trials.

From smallest radius to largest in order: 9 cm bar, 12 cm bar, 15 cm bar, 18 cm bar, 21 cm bar

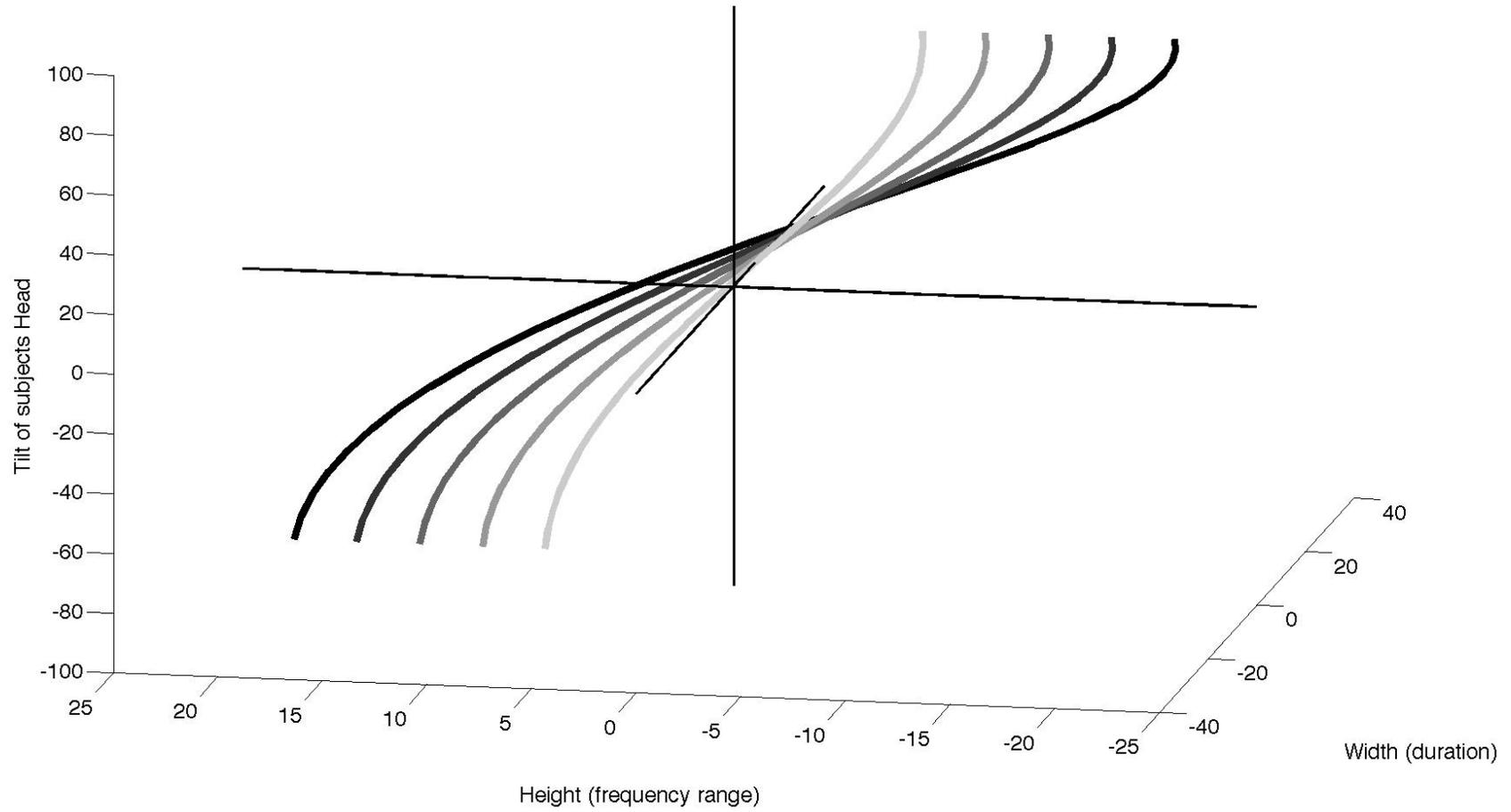


Figure 2.11. Head tilt and length constancy with a horizontal line. A horizontal line's dynamic change as participant tilts their head from vertical (no head tilt, tilt = 0) to 90 degrees left (negative tilt) or right (positive tilt). Each line represents a different line length ranging from 9 cm to 21 cm in length. Thus, in effect, learning aimed to identify all the data points within each curve as an "identical horizontal line," while discriminating across different curves as a "different length."

From smallest radius to largest in order: 9 cm bar, 12 cm bar, 15 cm bar, 18 cm bar, 21 cm bar

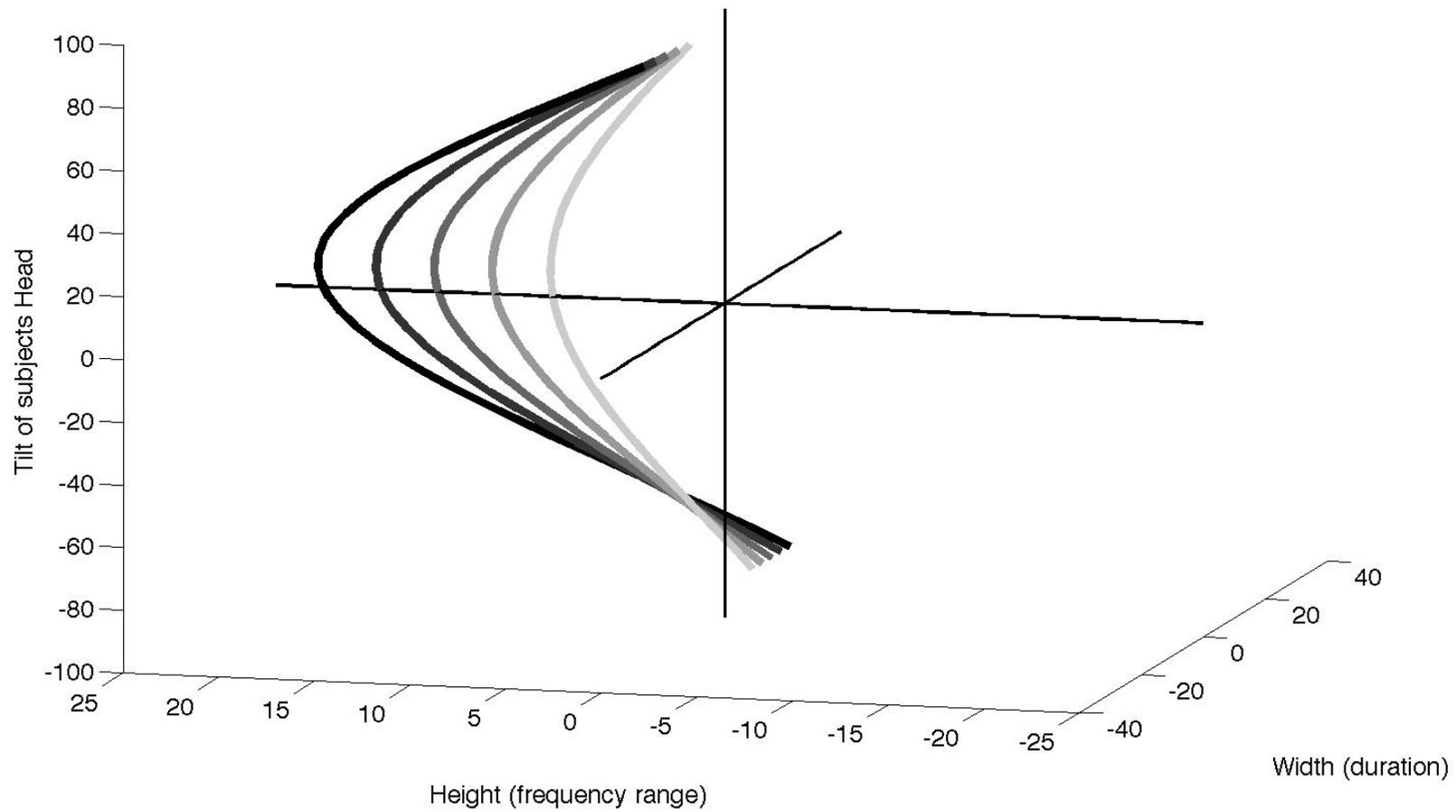


Figure 2.12. Head tilt and length constancy with a vertical line. A vertical line stimulus dynamically changes in horizontal width (duration of the sound) and vertical height (range of sound pitch) as the participant tilts their head from vertical (no head tilt, tilt = 0) to 90 degrees left (negative tilt) or right (positive tilt). Each line represents a different line length ranging from 9 cm to 21 cm in length. Thus, in effect, learning aimed to identify all the data points within each curve as an “identical vertical line,” while discriminating across different curves as a “different length.”

The enhanced learning due to head-tilt (retained in no-head-tilt trials) is consistent with a neural network using supervised training. In supervised learning, a neural network improves at classification with more training images and correct answer pairs presented¹⁷. In length constancy training, with no-head-tilt in each trial, the number of unique training images presented to the neural network is correlated with number of trials. But when the participant uses head-tilt, the number of unique training stimuli presented increases by a factor of at least *three*, one factor for each head tilt (vertical, left tilt, and right tilt), because each head position is a unique training stimulus. Each head position presents different parameters to calculate length, effectively tripling the training stimuli during supervised training and making classification more accurate (Changizi, Hsieh, Nijhawan, Kanai, & Shimojo, 2008). Further, while this calculation is based on snapshots at each head tilt, dynamic feedback would be even more “educational” to the network.

Active, as opposed to passive, interactions with the environment have been proven to be more effective for sensory-motor learning (Held & Hein, 1963). Some argue even more strongly that active interaction is crucial to visual awareness (O'Regan & Noe, 2001). Reynolds and Glenney showed that interactive two-participant training generated better performance at a vOICE device localization task than typical active or passive training with one participant (Reynolds & Glenney, 2012). J. Gibson's classical concepts such as “dynamic, direct perception” or “picking up higher-order invariance from the input affordance” may be relevant (Gibson, 1950b). Modern neural computation modals indicate that the brain is associative (Dayan, Abbott, & Abbott, 2001), using synapse weighting to correlate related properties between neurons that

“recognize” objects (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Reafferent signals or feedback from motor commands have been hypothesized by V. Mountcastle to provide a memory-based prediction to optimize sensory-motor learning (Mountcastle, 1978). Such re-afferent signals seem to critically enhance SS spatial perception and, therefore, constancy. More critically, the parametric analyses of the constancy with regard to the head tilt (Figure 2.11 and 2.12) not only reveal how the brain learns it with the SS device, but also in principle capture the learning of constancies in natural viewing. Thus, the implications of the current results/analyses go beyond just the SS learning.

Constancy learning with the vOICE demonstrates the dramatic plasticity of the adult brain. Tactile and auditory sensory substitution learning functionally recruits visual regions via extensive plasticity in blind and sighted users (Amedi, et al., 2007; Kupers, et al., 2010; Merabet, et al., 2009; Poirier, De Volder, & Scheiber, 2007). In particular, SS face stimuli activates the FFA (fusiform face area), SS shape discrimination activates the LOtv (Lateral Occipital tactile visual) area, and SS reading activates the VWFA (visual word form area)(Amedi, et al., 2007; Plaza, et al., 2009; Striem-Amit, et al., 2012). In addition, repetitive TMS had shown that congenital and late-blind users causally recruit visual regions for SS processing (Collignon, et al., 2007; Merabet, et al., 2009). Although no previous studies have brain-scanned the constancy learning with SS, a broad network of regions from sub-cortical auditory areas, primary auditory regions, multimodal regions and then visual regions may play a role. Further, active sensory feedback between these regions and motor areas likely improves multisensory network efficacy. If multisensory experiences and feedback shape such a neural network, active, as opposed to passive or static, learning procedures may enhance network shaping.

Length and orientation constancy are critical for rehabilitative use of SS, especially since it will enable “vision”-like processing and stimulus externalization. Constancies allow for the neural association of stimuli that have different proximal properties but represent the same object or feature in real 3D space, thus making a cohesive representation of external objects. SS devices have yet to aid the large population of blind people still limited in their daily-life functionality. Externalization of objects via length and orientation (and other) constancies could be the first critical stepping stone in the training process towards higher functionality at more complicated tasks in cluttered natural environments.

In sum, critical perceptual properties such as constancy and externalization can be achieved with current sensory substitution devices. Dynamic interaction with stimuli is shown to be critical to learning with sensory substitution owing to sensory-motor engagement and additional training information provided to the cognitive neural network.

CROSSMODAL INTRINSIC MAPPINGS MAKE
AUDITORY SENSORY SUBSTITUTION EFFORTLESS

Introduction

Sensory substitution studies have shown that sighted and blind participants can recognize and localize natural and artificial objects with sensory substitution given that participants have extensive training (one week to three months) and use top-down attention (Amedi, et al., 2007; Auvray, et al., 2007; Bach-y-Rita, et al., 1969; Bach-y-Rita, et al., 1998; Chebat, et al., 2011; Poirier, De Volder, & Scheiber, 2007; Proulx, et al., 2008). Whereas visual perception in the sighted is effortless and automatic, the usage of SS has so far been laborious, and this prevents devices from being successful commercially. No studies have investigated whether the processing of sensory substitution can be intuitive, or interpreted by entirely naïve participants with no device experience, training, or instruction. The only study that uses entirely naïve users is Auvray *et al.*, where they test whether distal perception (object perceived externally in perceptual space) can be learned without encoding knowledge of an auditory sensory substitution device, as detailed in Chapter 1 (p. 26) (Auvray, et al., 2005). It should also be noted that a SS visual acuity study used participants not trained with an SS device, but provided with a description of the device's vision-to-auditory encoding algorithm (Haigh, Brown, Meijer, & Proulx, 2013).

The current literature, reviewed in Chapter 1, seems to indicate that sensory substitution interpretation by trained users is a top-down cognitive process with attentive

concentration. Meanwhile, neural imaging studies on SS have so far shown the presence of plasticity, but uncertainty remains as to whether the plasticity is due to a top-down and attention-intensive process, or a bottom-up perceptual process (Amedi, et al., 2007; Poirier, De Volder, & Scheiber, 2007). Further, TMS studies have shown the visual activation from sensory substitution to be causally linked to task performance on the device in blind users (Collignon, et al., 2007; Merabet, et al., 2009). The current study (detailed in this chapter) is the first indication (among behavior or imaging studies) that sensory substitution interpretation (and potentially sensory substitution plasticity) does not always require top-down attention; rather it can rely on an automatic, bottom-up process.

Sensory substitution studies implicitly assume that blind or sighted participants cannot successfully interpret information provided by sensory substitution devices without both knowledge on the device encoding and sensorimotor training with it. However, the crossmodal correspondence literature (also called crossmodal associations, synaesthetic correspondences or associations, or intrinsic mappings) has shown that an intrinsic mapping exists between modalities (Spence, 2011). This intrinsic mapping may allow participants to perform tasks without any training, effort, or knowledge of the device encoding. For example, Figure 3.01 shows the intuitive matching of images to vOICe sounds by just using the amplitude modulation rate of the sound. The crossmodal mapping used in this example (amplitude modulation rate of sound to visual spatial frequency) is well-known, and has been studied in detail by Guzman-Martinez *et al.* (2012).

The crossmodal correspondences could further be used to enhance sensory substitution training by building on intuitive crossmodal features rather than ambiguous and unimodal visual features. Vision and audition correspondences can be generated by a common crossmodal feature, such as amplitude (brightness for vision, intensity for sound). On the other hand, seemingly unrelated modality-specific features have also been found to be matched, and matching can occur even at an abstract level (such as emotional response elicited) (Spence, 2011). It has been argued that these crossmodal mappings are learned priors within a Bayesian framework of crossmodal integration (Ernst, 2007). The encoding of vOICe is based on long-evidenced correspondences across vision and audition, such as the matching of brightness and loudness intensity (Stevens & Marks, 1965), spatial height and pitch height (Pratt, 1930), and scanning from left to right similar to reading written English. Therefore, participants with no knowledge about the vOICe device may in principle be able to use crossmodal correspondences to naïvely match images with their correct vOICe sounds. The device had been designed (either by chance or on purpose) for effortless usage, but somehow this advantage has not been explored. In addition to basic stimuli such as comparing lines of different angles encoded into sound with vOICe, our pilot observations suggest that other stimuli such as textures may have strong intrinsic crossmodal associations, and thus may also be correctly interpreted by naïve participants. This points to a possibility of a radical shift in SS training strategy. The vOICe device is particularly useful at encoding textures, as left-to-right scanning generates a dynamic beat that temporally plays out coarse-to-fine-grained spatial frequencies.

The naïve interpretation of vOICE would indicate that explicit instructions on the audiovisual vOICE encoding are not needed for vOICE interpretation. However, if the users can interpret vOICE without encoding instructions, this indicates that an intrinsic crossmodal mapping is utilized for interpretation, albeit implicitly. Therefore, the automaticity of the interpretation of vOICE naïvely depends on the automaticity of the crossmodal correspondences underlying that interpretation. Crossmodal correspondences can be automatic or require additional attention resources to interpret, depending on the type of mapping and task (Spence & Deroy, 2013). Chapter 1 discussed automaticity in vision, with an emphasis on visual distraction automaticity tests. Distraction tasks evaluate whether the stimuli in question is attention-load insensitive; this is one automaticity criterion. However, there are other criterion of automaticity, such as the “goal independence criterion,” “the non-conscious criterion,” and the “speed criterion” (Spence & Deroy, 2013). Spence and Deroy’s review of crossmodal mappings automaticity indicate that auditory visual correspondences have some evidence of being goal-directed (*i.e.*, not automatic), but in contrast are speeded in the Implicit Associations Test (*i.e.*, automatic) (Parise & Spence, 2012; Spence & Deroy, 2013). The experiments discussed in this chapter will use load-insensitivity criteria for automaticity of vOICE and the crossmodal correspondences therein. The load-insensitivity measure for automaticity will be tested with a distraction task in audition as well as in vision during the vOICE sound interpretation (detailed below). While there are no papers on load-insensitivity of crossmodal mappings, there are studies of load-insensitivity of crossmodal interactions.

Distraction dual task designs have been used in studying the impact of high attention load on crossmodal integration. Alsius and colleagues studied the processing of

auditory and visual speech integration in the McGurk Effect while participants performed a distraction task (Alsius, Navarra, Campbell, & Soto-Faraco, 2005). Results indicated that reduced attention resources limited the McGurk effect. A study performed by Eramudugolla *et al.* indicated that ventriloquist aftereffect can occur under attention load, but that it is modulated by attention load (Eramudugolla, Kamke, Soto-Faraco, & Mattingley, 2011). Helbig and Ernst demonstrated that the weighting of visual and haptic stimuli is independent of attention load (Helbig & Ernst, 2008). These mixed multimodal results on attention load indicate that crossmodal mappings may or may not be independent of attention load. We will study this further in this chapter in application to the crossmodal mappings used in the vOICe device.

We address two crossmodal mapping problems in this Chapter: the engineering issue of optimally encoding vision into audition ($V \Rightarrow A$), and the psychological/neural decoding of SS via crossmodal correspondences ($A \Rightarrow V$). We began by studying the psychological/neural decoding of SS with the existing vOICe device encoding, to determine if vOICe can be intuitive. The results then suggested optimal methods for the encoding of vision into audition. In other words, once we know what works in vOICe, we can then accentuate those characteristics to make even more intuitive device encodings and training procedures

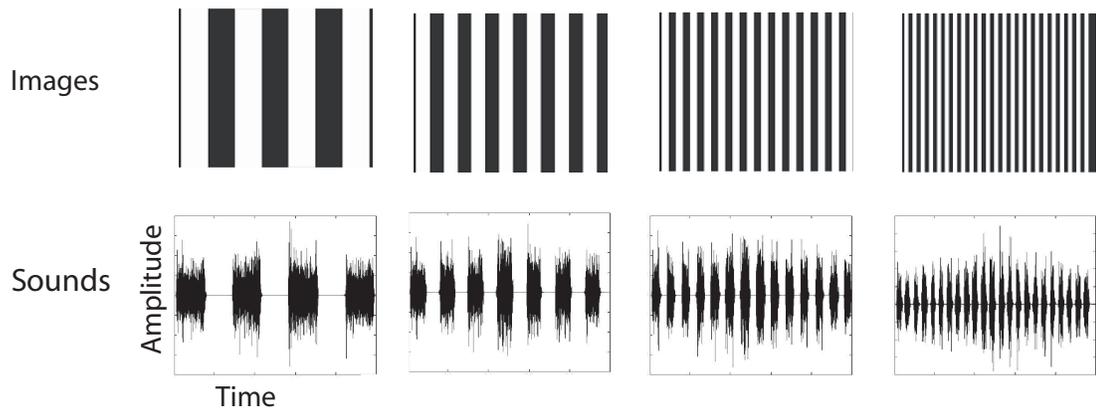


Figure 3.01. Example of intuitive image and vOICe matching. Figure 3.01 shows the example output from the vOICe (row 2) for a given set of images (row 1) used in bimodal matching experiments. Each row in the graphic is a different representation of the set of images: the first row is the visual representation, the second row uses just amplitude of the vOICe sound as a function of time to represent the image. Each column represents the same image or information. It is clear with this particular set of images and vOICe sounds that they have similar structure, and therefore are intuitive to match. In fact, it is clear that it is easy to match the images and sounds even if the positions of images and sounds were jumbled.

We hypothesize that textures will be intuitive with vOICE. Textures have been studied in detail in vision, and are an important element of monocular depth perception, visual segmentation, and automatic visual search (Palmer, 1999). Cues in monocular depth perception such as texture gradient (texture elements become smaller with distance) and texture accretion and deletion (texture elements disappear and reappear with lateral movement) are important elements of monocular depth. In visual search, unique texture elements can be identified in either a parallel or serial manner (Bergen & Julesz, 1983). With parallel search, the unique element pops out and can be identified at the same speed independent of the number of distractors. In the serial search, the unique element localization depends on the number distractors (no pop-out). Textures can also be used in vision for segmenting a scene into different objects and/or visual regions, and can be used in object shape identification (via distortion of texture elements). As an important and prevalent element of vision it is logical that textures would also be valuable to the processing of vOICE stimuli.

Methods

The role of crossmodal correspondences was tested with naïve ($N = 5-7$) and trained sighted ($N = 4$) participants in a bimodal matching task (Figure 3.02). First, all stimuli were presented as a preview (all three to four images and then associated vOICE sounds in random order), and then participants heard one sound and were asked to match one of three presented images to the sound (3AFC). Naïve participants were not told the vOICE encoding scheme, nor that the sounds were from the vOICE. Participants were asked to match the image and sound that carried the same information; if uncertain, participants were told to guess. Feedback on performance was not provided to

participants. Images were compared in sets of three or four so that particular image features and types could be tested separately. Image types ranged from natural to artificial images, and from simple to complex images. Images sets included vertical bar textures of different thickness, circular patterns of different element sizes, and images of natural textures (Figure 3.06). All images were presented in grayscale, as vOICe sounds do not convey color information. A total of 24 image sets were tested (all images are included in the supplementary materials). The naïve sighted participants are different participants from the naïve trained participants.

The crossmodal mappings underlying vOICe's interpretation were tested on naïve sighted participants ($N = 8$). Participants performed a bimodal matching task of the same design as the original (detailed above), but with different encoding schemes to test the value of different crossmodal mappings. Different encoding paradigms were generated by altering the images inputted into the vOICe encoding software (for example: The inverted coding of dark regions louder than bright regions was generated by inverting image brightness before inputting the image into the vOICe software). The encoding inversions tested (on top of original; [0]) were: (1) dark regions louder than bright regions, (2) scanning right to left, (3) high frequency on the bottom, and, (4) scanning top to bottom and high frequency on the right (Figure 1.4 has the original vOICe encoding). The order of testing the different encoding inversions on participants was randomized (including the original mapping). All participants completed all five of the different encoding types (four inversions and one original) in one session.

Automaticity of vOICe interpretation via an attention load experiment was tested with a dual task design. In the first experiment, participants counted backward in 7s from

a random number displayed (between 100 and 112), while counting the vOICe sound played (vOICe sound started 10 seconds after counting started) ($N = 8$) (Figure 3.03). Participants then matched the vOICe sound to one image of three images displayed (3AFC, same design and image specifications as the bimodal matching experiment). The same participants also performed the original bimodal experiment (*i.e.*, with no counting) in the same session, which was used for comparison ($N = 8$)(original encoding, *i.e.*, “0” in above list). A subset of the same participants performed a visual search distraction task in a second session ($N = 6$; randomly chosen from the 8 participants above) (Figure 3.04). These participants searched for an F within 50 E’s randomly placed in a 100-by-100 location grid in a single image. The E and F locations were jittered vertically and horizontally by up to 50 pixels. The F was present in half of trials, and absent in half. The image to be searched was presented on screen until participants responded to the visual search question. The visual search image was 10 inches by 10 inches, and each letter was 0.25 inches by 0.5 inches on screen. Participants sat about 25 inches from the 27 inch iMac screen where the images were presented. The vOICe sound played at the beginning of the visual search task. The participant was encouraged to continue searching while the sound was played. Participants then matched the vOICe sound to one image of three images displayed (3AFC, same design and image specifications as the bimodal matching experiment).

The tactile auditory mappings were tested via a bimodal matching task (Naïve sighted $N = 2$, Naïve blind $N = 2$, Trained blind $N = 2$ (both late blind)) (Figure 3.05). The set of the experiment was similar to the visual auditory bimodal matching. First, three to four tactile patterns (4 inches by 3.25 inches) were explored and the associated

vOICe sounds were played in random order, as a preview. Then, participants listened to one of the vOICe sounds and matched it to one of three tactile patterns presented on a desk surface (3AFC). Participants were asked to match the image and tactile pattern that carried the same information. The tactile-auditory matching task instructions were read aloud to the blind or blindfolded sighted by the experimenter and the participant's responses (conveyed orally) were inputted by experimenter. Tactile stimuli were placed in front of the participants on a desk surface for exploration by the participant. Tactile patterns used were generated from black and white images containing two brightness levels, by adhering cardstock to the white regions, thereby raising them relative to the black by about 1 millimeter. Images of all tactile relief patterns are presented in Figure 3.11. The trained blind participants are the same participants as the naïve blind participants.

Sighted naïve participants also performed a vOICe memory task (mimicking the vOICe training tasks) ($N = 4$) for a between group comparison. Initially, the sounds from vOICe were played in random order twice, and a label (1-4) was given to each of the sounds. Then, in each trial, one of the sounds would play again and the participant would respond with the number that matches that sound. This memory task was performed on the same sets of images that were used for the bimodal matching task.

Participants performed all tasks at a 27-inch iMac computer station with Sony noise-cancelling headphones (MDR-NC7), and inputting responses into a keyboard. Psychophysics Toolbox and MATLAB were used to code the presentation of instructions and stimuli as well as recording responses. Images were presented in black and white on the iMac screen (image size: 4 inches by 3.25 inches) approximately 25 inches away

from the seated participant. Images were encoded into vOICe sounds using vOICe software from seeingwithsound.com using a 1 Hz scan rate. Screen brightness and audio loudness was set to be comfortable to the participant. Images used were retrieved on the internet or generated by experimenter in Adobe Illustrator. Images retrieved from the internet were occasionally modified in Adobe Illustrator or Adobe Photoshop.

All trained participants were trained for 8 days on the vOICe device on basic object localization and recognition as well as two constancy tasks (rotation and shape constancy). For more details, see Appendix B and Chapter 2 Methods, (p. 62-65). The vOICe device used a camera embedded in a pair of sunglasses or a webcam attached externally to glasses. Sighted participants were requested to close eyes during training and evaluation, and wore opaque glasses and/or mask. A camera provided live video feed of the environment, and we used a small portable computer to encode the video into sound in real time.

Complexity quantification was performed in MATLAB. Images were filtered with the Laplacian of Gaussian method (edge function) and then averaged to a single number per image that was averaged across an image set. The resulting number was correlated with the bimodal audiovisual matching performance.

ANOCOVA and correlation analyses were performed in MATLAB using the aocool and corr functions.

Results

In the original bimodal matching task (matching images to sound with vOICe encoding), naïve sighted participants ($N = 5$ to 7 participants, varied across stimulus sets)

performed significantly above chance (*i.e.*, $p < 0.05$) in 12 of 24 image sets tested, and trained sighted participants ($N = 4$) in 16 of 24 image sets (See Figure 3.06 and Appendix 1; Appendix 1 includes all images tested). Even with the strict Bonferroni multiple comparisons correction (*i.e.*, $p < 0.0021$), 5 of 24 image sets were above chance for naïve, and 8 out of 24 for trained.

The image sets tested can be divided into three groups: Artificial images (generally simple and generated by myself; Appendix A, Table A and B), non-modified natural stimuli (such as flowers, forests, natural textures; Appendix A, Table A and B), texture interfaces (natural textures artificially combined to generate interfaces; Appendix A, Table C). In the artificial stimuli, 6 out of 9 image sets (67%) are significantly above chance (*i.e.*, $p < 0.05$) for naïve sighted and 7 (78%) for trained sighted. If just non-modified natural stimuli are counted, of 7 image sets, 2 image sets (29%) were significantly above chance (*i.e.*, $p < 0.05$) for the naïve sighted, and 5 image sets (71%) for the trained sighted. Finally, for the texture interface group, 4 of 8 image sets (50%) are significantly different from chance (*i.e.*, $p < 0.05$) for the trained and naïve. Therefore, the artificial stimuli seem to be the strongest group for matching images and sounds in both naïve and trained, likely due in part to their simplicity (for example: A single line or dot on a black background).

When the naïve and trained are compared directly, only in 1 image set out of 24 was the naïve performance significantly different from the trained performance (row 1 of Table C in Appendix A, $p < 0.01$). The image set is a set of texture interfaces for jeans and wood floor texture. It is useful to note that this image set for naïve vs. trained does not survive the Bonferroni multiple comparisons correction (*i.e.*, $p < 0.0021$). When the

results for each image set are averaged across naïve participants and then trained participants, these averages were found not to be significantly different for the naïve vs. trained participant groups ($p < 0.30$). Therefore, surprisingly, the naïve and trained groups are quite similar in their bimodal matching performance.

It was an unexpected result that natural stimuli could be intuitive to interpret with sensory substitution. Natural stimuli (such as a natural texture) have more spatial frequencies and brightness variation than the typical simplified lab image (a vertical line, for example). Most participants being trained on sensory substitution as reported in the literature begin with a simplified lab environment, such as a white isolated object on black felt background, and only experience a natural environment with the device after at least several training sessions. Our study indicates that this approach to training could be flawed. We have found that some natural stimuli (such as natural textures) are rich in crossmodal correspondences, and therefore are easy to interpret with vOICE. It might be better to begin training participants with a crossmodal correspondence-rich environment that includes both natural texture tasks and the simplified lab tasks.

Crossmodal mappings underlie the vOICE encoding intuitiveness. While this is a logical conclusion from the results in Figure 3.06, it is not explicitly proven that crossmodal mappings are the critical element that makes vOICE understandable to the entirely naïve. Further, it is unclear which mapping within the vOICE encoding is the most important for accurate interpretation. To address these issues, we reversed each of the primary vOICE encodings or crossmodal mappings, and then tested the new reversals in comparison to the original vOICE encoding. If the encoding or crossmodal mapping reversal significantly reduces the participants' accuracy at matching images and sounds,

then that mapping is important to correctly naïvely interpret vOICe.

Results from 8 sighted naïve participants (in Figure 3.07) indicate that two crossmodal correspondence inversions have a significantly reduced accuracy compared to the original encoding. The correlation of brightness and loudness was significantly less accurate when reversed for two most real-world-like image sets: Interfaces ($p < 0.00$) and Natural Textures ($p < 0.04$) (second and third image sets in Figure 3.07). The XY orientation of the encoding (scanning left to right, and high pitch at the top of the image) was also significantly less accurate when reversed (scanning top to bottom, and high pitch on right of image) for one image set: Bars of different thickness ($p < 0.00$) (first image set in Figure 3.07). When all the images are summed together, both the mapping of brightness and loudness ($p < 0.01$) and XY orientation ($p < 0.00$) when inverted had significantly less accurate performance than the original encoding (Figure 3.08).

The implications of the crossmodal mapping tests are that two encoding elements are particularly important to image interpretation with vOICe: Brightness correlating with loudness, and the XY orientation of the encoding (*i.e.*, the scanning from left to right rather than top to bottom, and high pitch with the top of the image rather than the right). It appears that the reversal of the encoding from top to the bottom or from the left to the right can be tolerated, but the switching of the Y and X axis encodings is problematic to interpretation. The problem of switching Y and X axis encodings further emphasizes the anisotropy of the vOICe encoding (unlike vision) and the importance of displaying information on the X axis, where the highest resolution occurs (rather than the Y axis). In particular, the images that test well with vOICe have information displayed horizontally, and when the XY encoding is switched, the information in the X direction is

less detectable by the lower Y -axis encoding resolution, thereby reducing accuracy. The value of brightness correlating with loudness makes sense, as most bright objects in a dark area are the most interesting (rather than vice versa). However, its value is also fortified by the auditory system's acute ability to recognize the presence of sounds, and its inability to recognize the absence of sounds. Therefore, the combination of these two facts makes the brightness translation to loudness highlight the most important image elements (*i.e.*, the bright elements), whereas the reverse encoding (darkness translates to loudness) obscures the most important elements.

The interpretation of vOICe does not require explicit knowledge of the sound-to-image encoding; however, this doesn't fully prove that vOICe interpretation is effortless. The vOICe interpretation relies upon crossmodal correspondences (as highlighted in the previous experiments), and crossmodal mapping interpretation can be automatic or require attention (discussed in Chapter introduction). Therefore, the automaticity was tested for naïve interpretation of vOICe sounds with an attention distraction experiment. The audio distraction task used for vOICe was counting backward in sevens while the vOICe sound was played (experiment detailed in methods). The visual distraction task was a visual search task, where participants searched for an F within 50 E's. The dual task matching accuracy (both audio and visual) was not significantly different from the original vOICe bimodal matching task for any of the 4 image sets tested (Figure 3.09) ($N = 8$). When the data are summed across image sets, the visual and auditory distraction task accuracy were both still not significantly different from the original bimodal matching task (auditory distractor: $p < 0.08$, visual distractor: $p < 0.31$). Therefore, this result shows that naïve vOICe interpretation is independent of attention load. This fulfills

one important criterion of automaticity and indicates that naïve vOICe interpretation is effortless in at least one measure.

Does image complexity matter to the untrained participants' performance? To examine this, we defined image complexity by an edge metric that quantifies the number of vertical and horizontal edges. The trained and naïve sighted participant performance both weakly anti-correlated with complexity, as measured by the edge metric (Naïve participants: $\rho = -0.3491$ $p < 0.09$; Trained participants: $\rho = -0.3858$, $p < 0.06$) (Figure 3.10). This result indicates that complexity may make images less intuitive to interpret. However, more importantly, a linear fit to the data indicated a performance above chance at even large complexity values for the naïve and trained participants. The trained and naïve anti-correlations with complexity had slopes and intercepts that were not significantly different from each other (ANOCOVA analysis, $p_{slope} < 0.73$, $p_{intercept} < 0.27$). It is likely that “complexity” can partially mask the crossmodal correspondences or dilute the crossmodally relevant information with unimodal noise. Nonetheless, some of the more “complex” stimuli such as natural textures revealed way-above chance performance that is likely due to direct selection of a high density of crossmodal mappings (such as coarse to fine spatial frequencies) (Figure 3.02 and Appendix A).

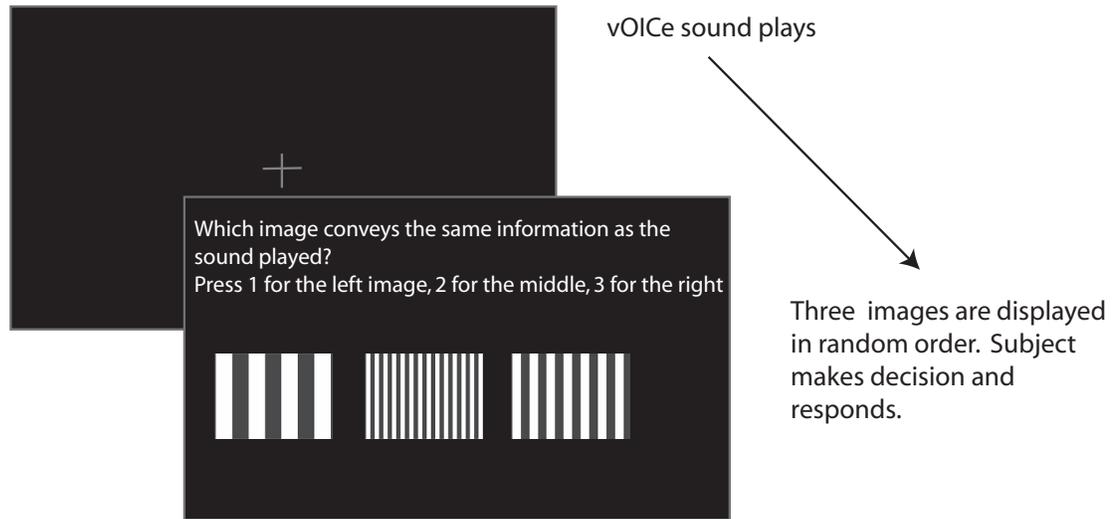


Figure 3.02. Experiment design for visual-auditory matching. As detailed in methods, participants performed matching the images and vOICe sounds while at a computer. First a vOICe sound would play, and then participants would be required to choose an image that seemed to match that sound the best, or contained the same information. Sighted participants responded by inputting a number into the keyboard: 1 for the left image, 2 for the middle image and 3 for the right image.

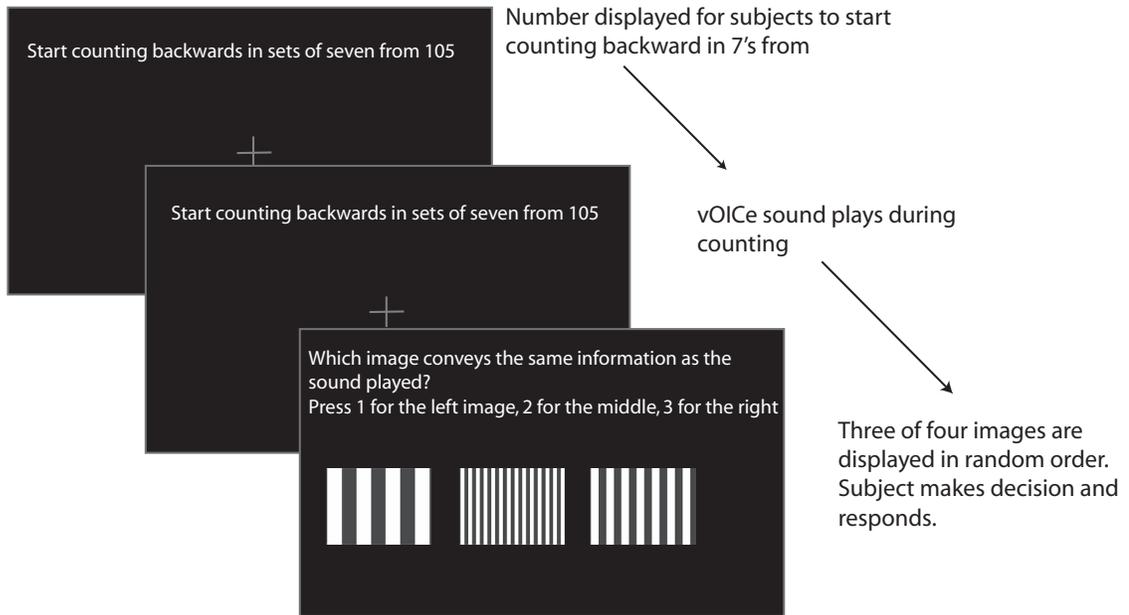


Figure 3.03. Experiment design for auditory distraction during visual-auditory matching. During the auditory distraction version of the auditory-visual matching of images to vOICe, participants were distracted by counting backward in sets of seven. The experiment was designed such that participants count backwards (beginning with the number presented on the screen), and during counting a vOICe sound plays. The final task is for the participants to match the sound heard while counting to one of the three images presented. Participants responded by inputting to the keyboard: 1 for the left image, 2 for the middle image, and 3 for the right image.

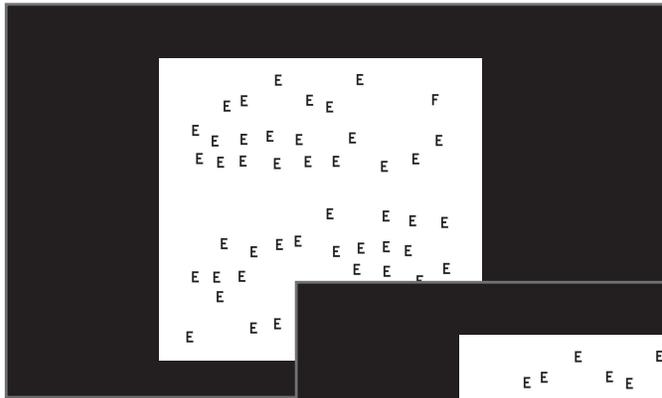
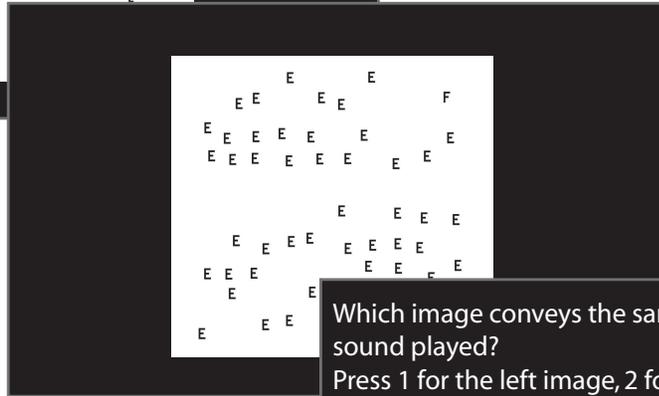


Image of E's and F's is displayed for subject to start searching for the F. vOICe sound plays.



Subject responds if an F is present

Which image conveys the same information as the sound played?
Press 1 for the left image, 2 for the middle, 3 for the right

The first set consists of 6 bars of medium width with wide spacing. The second set consists of 12 bars of narrow width with narrow spacing. The third set consists of 8 bars of medium width with medium spacing.

Three of four images are displayed in random order. Subject makes decision and responds.

Figure 3.04. Experiment design for visual distraction during visual-auditory matching. During the visual distraction version of the auditory-visual matching of images to vOICe sounds, participants were distracted by searching for an F within a field of 50 E's. While searching for the F, a vOICe sound is played. The participants finished the searching task by inputting to the keyboard 1 if an F is present, and 2 if an F is absent. The second task then appears, wherein the participants are required to match the vOICe sound played while searching to one of three images presented. To complete the matching task, participants input to the keyboard 1 for the left image, 2 for the middle image, and 3 for the right image.

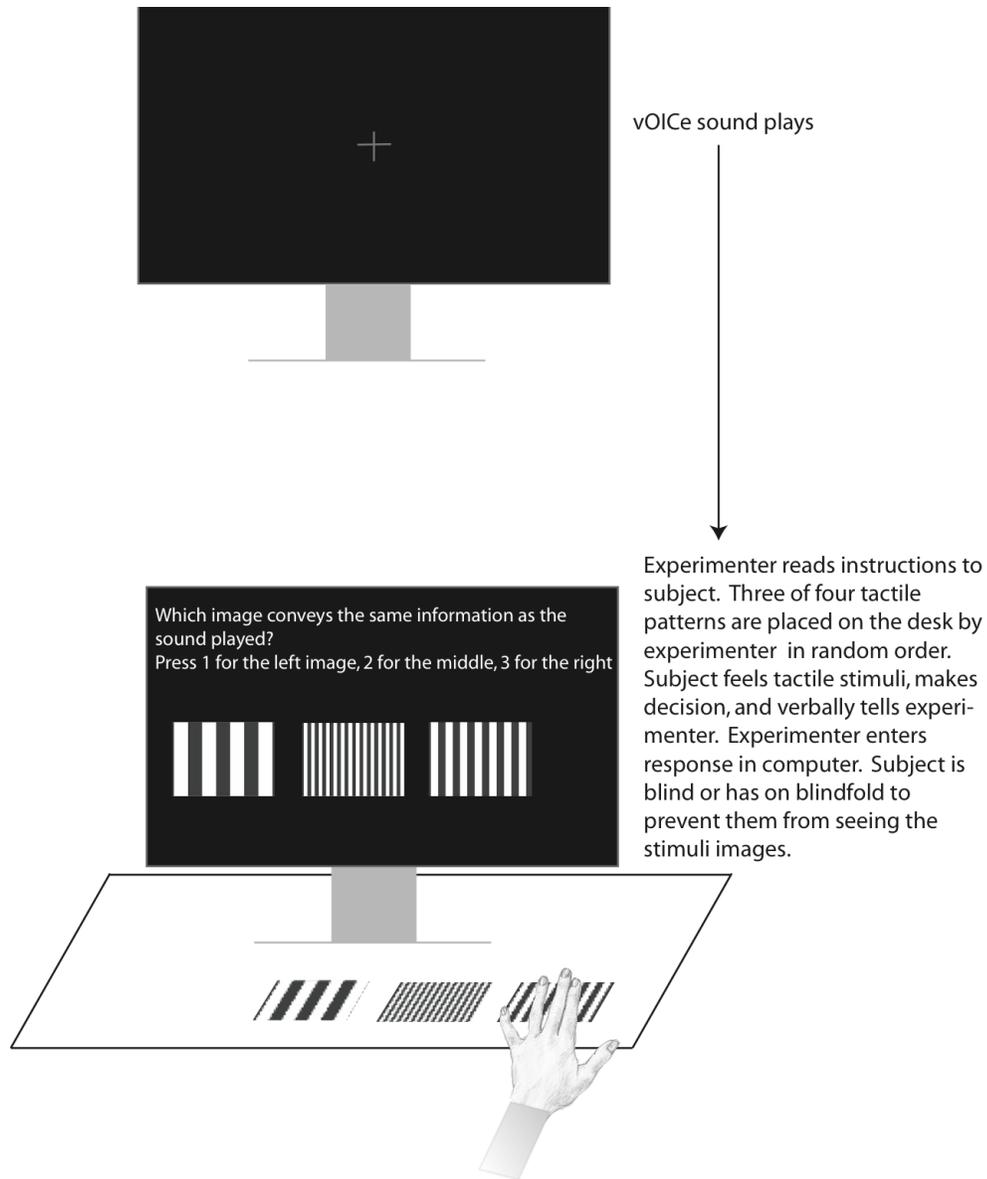


Figure 3.05. Experimental design for tactile-visual matching. Blind and blindfolded sighted participants were read the instructions for the task by the experimenter. The task began with a vOICe sound playing in headphones; then, three tactile patterns would be placed in front of the participant for tactile exploration. The participant indicates the chosen pattern, and the experimenter enters the corresponding number in the computer.

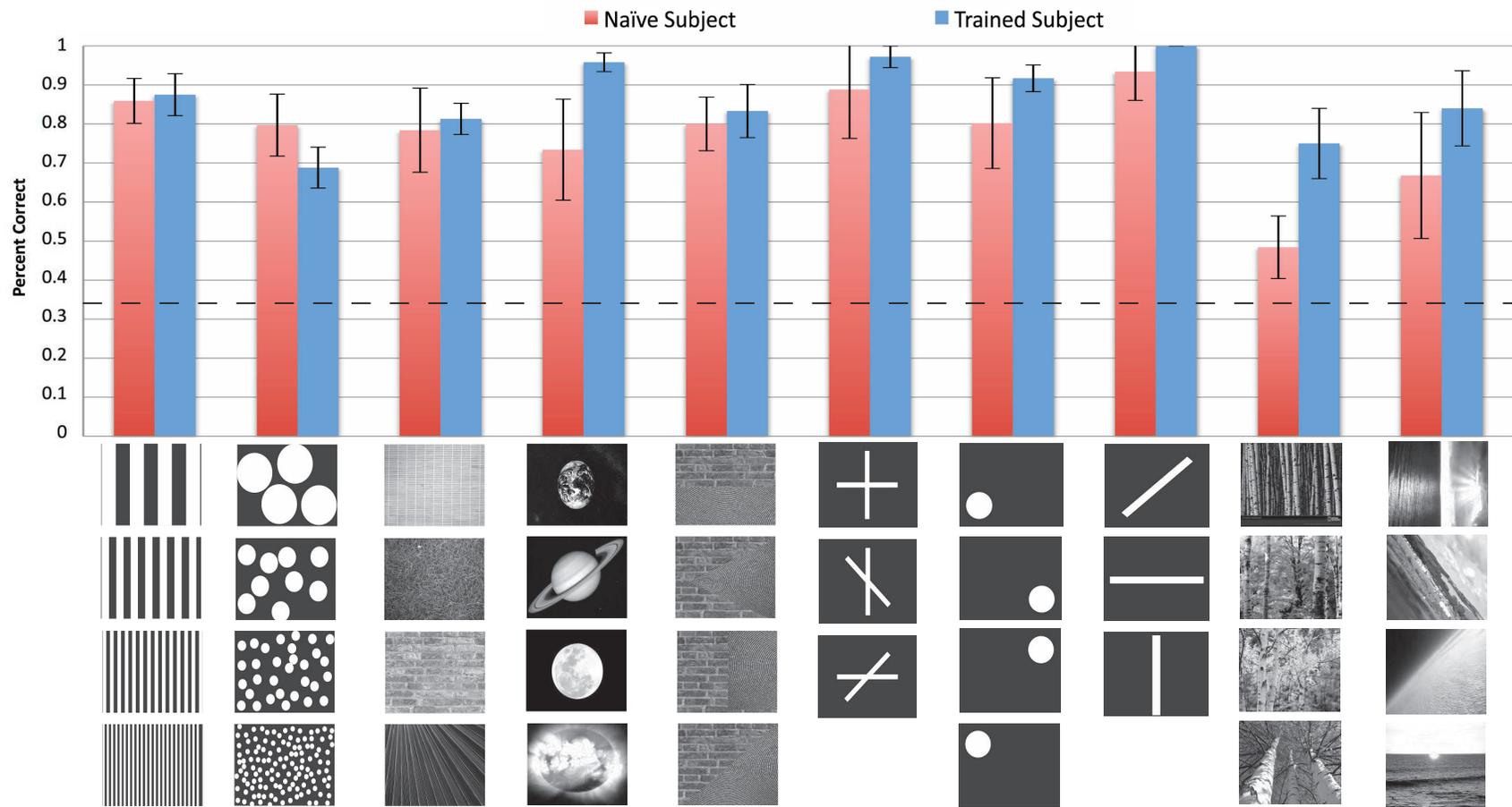


Figure 3.06. Select vOICe data and images. Data and images from a select set of images encoded into vOICe sounds and tested on naïve and trained sighted participants. Participants were tested at matching a vOICe sound to the corresponding image out of three presented. The error bars are the standard deviation across participants. All data presented in Figure 2B is significantly different from chance ($p < 0.05$), except the naïve percent correct for the last two image sets on the right (*i.e.*, trees and horizon images).

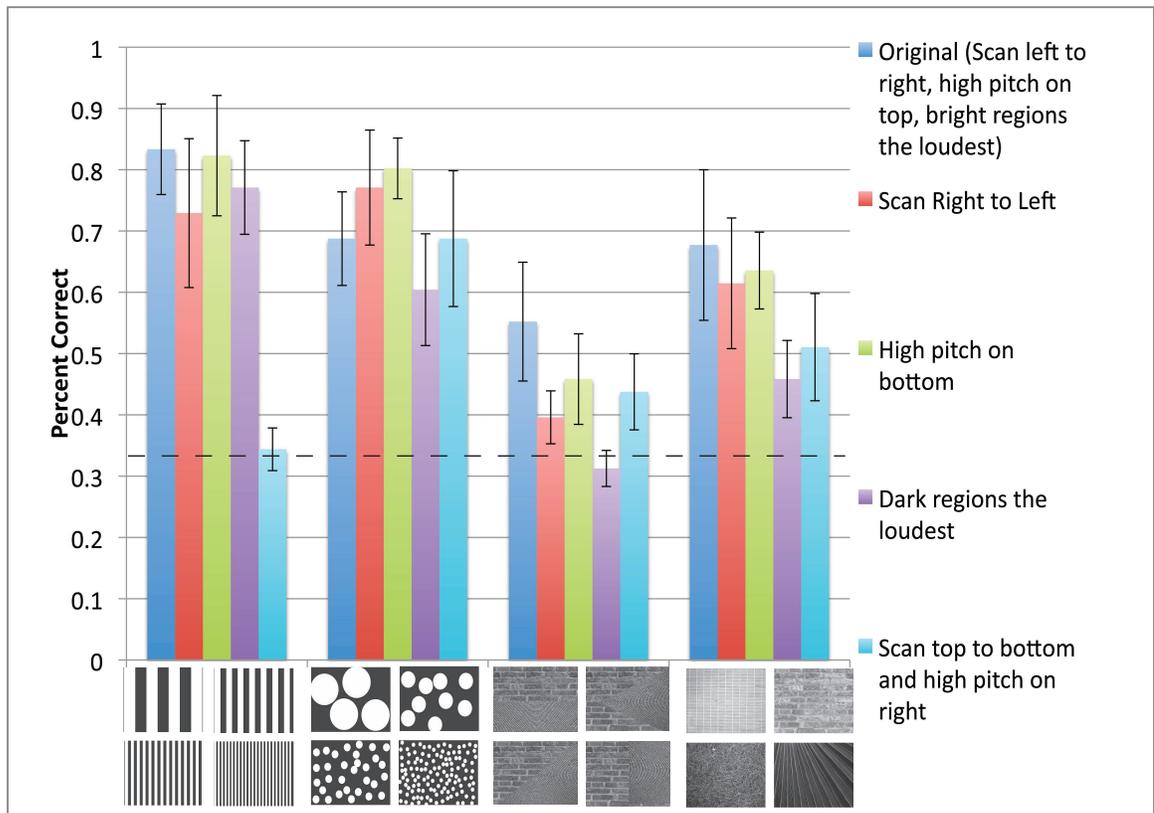


Figure 3.07. Tests of vOICE crossmodal mappings. Modifications in the vOICE auditory to visual mapping were tested with naïve participants to determine each of the crossmodal mappings' importance. The error bars are the standard deviation. The dashed line is chance.

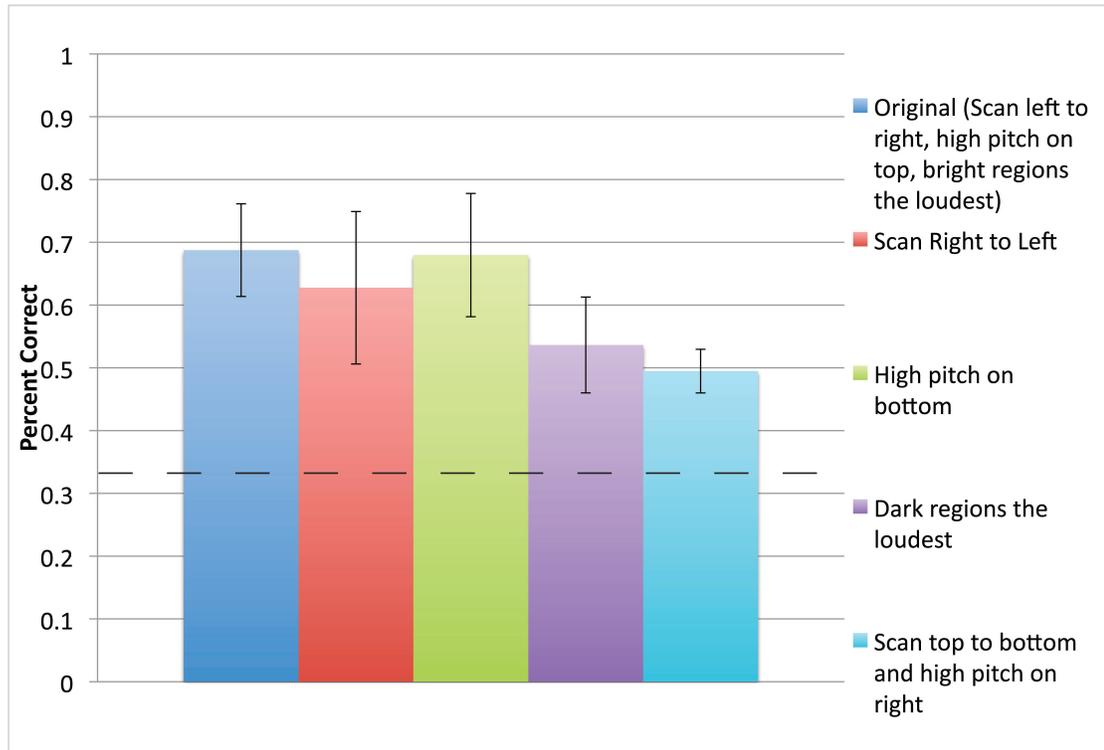


Figure 3.08. Tests of vOICE crossmodal mappings summed across images. Modifications in the vOICE auditory to visual mapping were tested with naïve participants to determine each of the crossmodal mappings' importance. The images sets were averaged together to generate a generalized percent correct for all four image sets tested (Figure 3.07 shows individual image set data). The error bars are the standard deviation. The dashed line is chance.

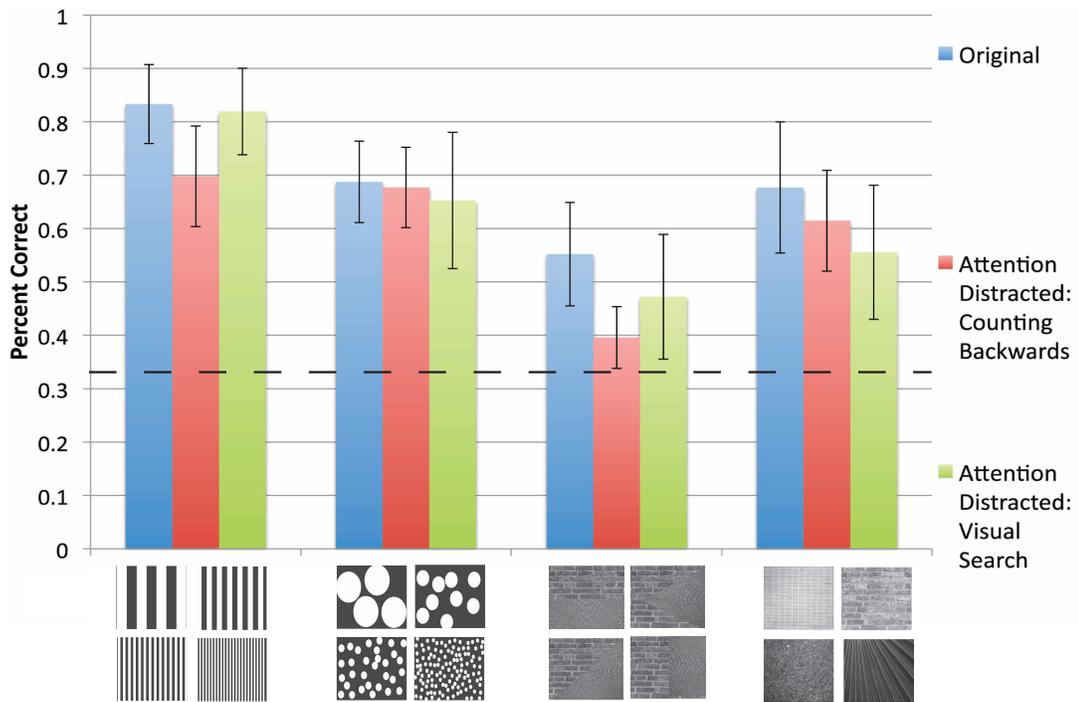


Figure 3.09. Auditory and visual attention distraction vOICe data. Naïve (untrained, and no encoding knowledge) participants matched vOICe sounds with images while performing a distraction task (either counting backward in sets of 7 from a random number [$N = 8$] or visual search [$N = 6$]). Participants then matched the sound heard to 1 of 3 images displayed. The attention distraction data is compared to the original matching of sounds to images without distraction in the same participants. Error bars are the standard deviation, and the dashed line is chance.

The naïve sighted participants can perform marvelously well matching visual images to sounds, but the real question relevant to sensory substitution should be whether the same (multimodal mappings) can be applied to, say, auditory and tactile modalities in naïve blind participants. Thus, we tested blind participants on matching sounds to tactile (relief) patterns that corresponded to the visual patterns described above for lines of different thicknesses and circle patterns of different sizes, and they also performed above chance (Figure 3.11, Bars of different thickness: Late Blind Naïve ($N = 2$) 50%, Late Blind Trained ($N = 2$) 71%, Sighted Naïve ($N = 2$) 67%; Dots of different sizes: Late Blind Naïve ($N = 2$) 50%, Late Blind Trained ($N = 2$) 71%, Sighted Naïve ($N = 2$) 63%; Chance 33%). Although the late-blind data for tactile-auditory matching is weaker than the sighted data for auditory-visual matching, the late-blind will also likely have a hidden and untestable vision-audition intrinsic mapping from past visual experience that does not appear on the tactile-audition matching test performance. Such a hidden visual-auditory mapping may assist or facilitate in the learning of vOICE by the late blind.

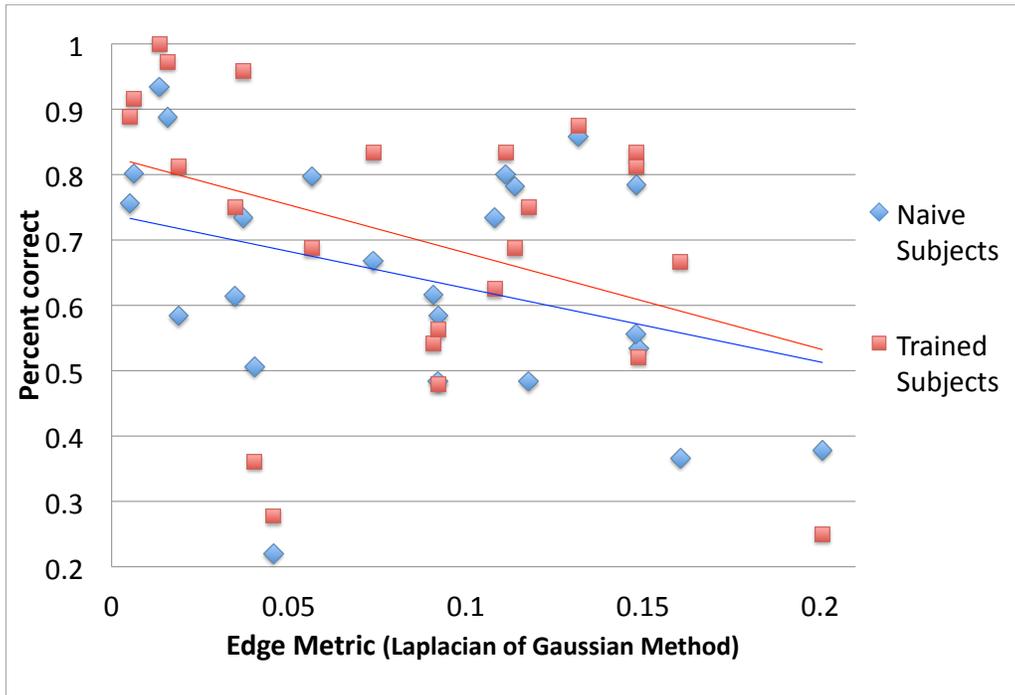


Figure 3.10. Correlation between bimodal matching data and edge metric. Correlation data: Naïve Participants: $\rho = -0.3491$, $p < 0.09$; Trained Participants: $\rho = -0.3858$, $p < 0.06$. Edge metric calculated in MATLAB by filtering images for edges and then averaging all pixels.

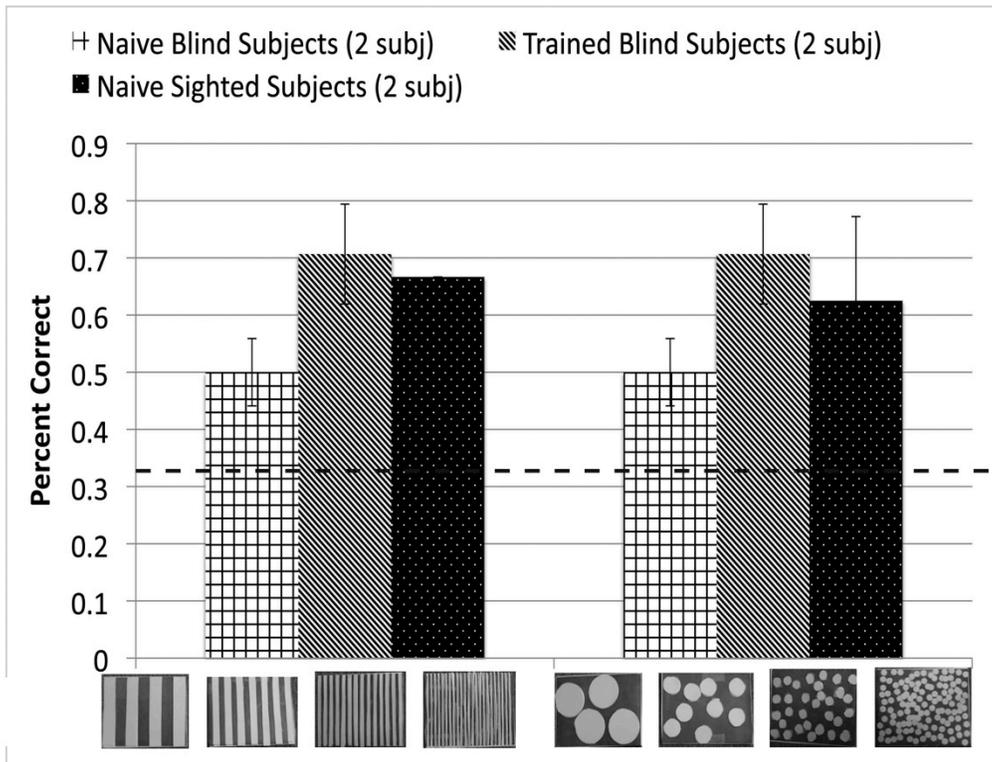


Figure 3.11. Data and images from matching of vOICE sound and tactile patterns. The tactile patterns are derived from image textures previously tested. Participants were tested at matching a vOICE sound to the corresponding tactile pattern out of three presented. The error bars are the standard deviation across participants. The white regions of the tactile patterns are raised relative to the black regions.

The matching experiments demonstrated that participants have the ability to crossmodally match vOICe sounds and images. It was yet unclear whether this crossmodal ability affects more conventional, unimodal (*i.e.*, just auditory) training with the device. To demonstrate the relationship between vOICe training and crossmodal matching ability, naïve sighted participants also performed a memory task with the same stimuli as in the bimodal matching task (detailed above). Participants were told a label (1-4) to remember for each sound, and then asked to recall the label when a random one of vOICe sounds was played. The memory task format is similar to most sensory substitution training tasks. There, participants are presented with an object or stimulus and allowed to explore or listen to it, and then told a label such as “pencil” or “square.” The participant would be asked later whether they could identify the objects when presented in random order. Such a memory-based label task is in the same format as our memory task with the intuitive sensory substitution stimuli. Participant performance on this auditory memory task (chance: 25 percent) correlated significantly with the performance on the crossmodal matching task (chance: 33 percent) with a *rho* of 0.7139 ($p < 8.8 \times 10^{-4}$) (Figure 3.12). The result therefore indicates that the participants’ ability to remember and interpret sensory substitution stimuli correlates significantly with the density of crossmodal mappings (as measured by our crossmodal matching task). Therefore, crossmodal intrinsic mappings provide a common basis for sensory substitution training as well as adaptive behavior and scene perception in the real world with the device. Crossmodal correspondences are the unrecognized common key to the relative intuitiveness/ease of existing vOICe training tasks.

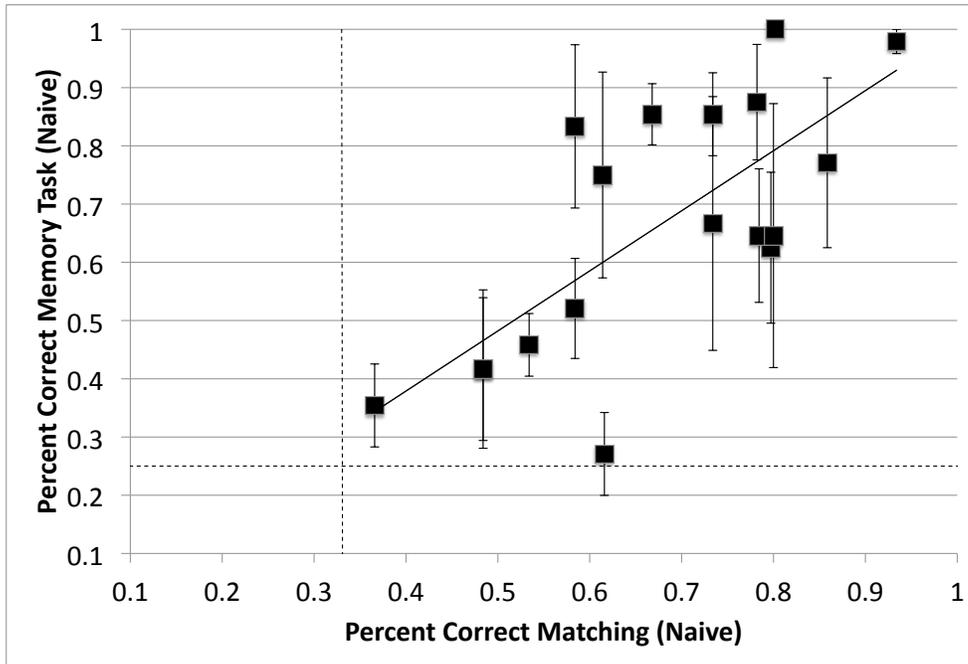


Figure 3.12. Correlation between the bimodal and unimodal tasks. In the bimodal matching task, the participant matches vOICE sounds to images, and in the unimodal memory task, the participant indicates the remembered label for each vOICE sound. The memory task is the same as most vOICE training tasks. Dashed lines are chance for each of the tasks.

Discussion

Sensory substitution training has a hidden assumption that the primitives of sensory substitution perception will be the same as the primitives of vision, such as dots, lines and intersections. While sensory substitution is vision-like, it may have crossmodally intuitive primitives that are different from the classical visual primitives, and should not be overlooked. Training protocols that are specially designed to access intrinsic mappings as primitives may enable faster training and more ease of use. If intuitive stimuli such as textures are the starting point of vOICe training, followed by the gradual increase of image complexity (but also closer to the real-world), participants may be able to learn to use devices more effectively and effortlessly with a shorter training period. Training could also use image-processing filters to heighten textures in the natural images (such as a high pass filter), thereby making them more intuitive. Note that this is a grossly different approach from the conventional (more effort-demanding) training, where trainees are forced to learn geometric primitives and then more natural cluttered scenes constructed from these primitives.

This study indicates that participants can interpret vOICe stimuli with no knowledge of the audiovisual encoding. The strongest crossmodal correspondences that underlie this naïve vOICe interpretation were found to be brightness to loudness mapping and the *XY* mapping orientation. Finally, the naïve interpretation of vOICe was shown to be automatic (attentional load insensitive) with a dual task design.

Sensory substitution interpretation and functional ability is generated by multimodal interaction and crossmodal plasticity. Crossmodal mappings are the foundation of sensory substitution interpretation, and if used intelligently in device

training and design, could dramatically improve functional outcomes. The fundamental bottleneck towards a commercial product may be removed by vigorous crossmodal plasticity kick-started from such an advantageous start point.

AUTOMATICITY OF VOICE CROSSMODAL PLASTICITY

Introduction

As detailed in Chapter 1, participants trained to interpret sensory substitution (SS) sounds have crossmodal plasticity generating visual activation in response to SS sounds (Amedi, et al., 2007; Arno, et al., 2001; Kupers, et al., 2010; Merabet, et al., 2009; Poirier, De Volder, Tranduy, et al., 2007; Poirier, De Volder, & Scheiber, 2007; Poirier, De Volder, et al., 2006; Ptito, et al., 2005; Renier, Collignon, Poirier, Tranduy, Vanlierde, Bol, Veraart, & Devolder, 2005; Renier & De Volder, 2010). The vOICe and SS in general have many of the characteristics of vision such as depth illusions, recognition, localization, and processing of SS stimuli in early visual areas. Nevertheless, unlike vision, vOICe interpretation is slow and laborious even after extensive training, and therefore is often assumed to be processed top-down (involves cognition). In contrast to vOICe, vision is often perceptual and passive (*i.e.*, automatically occurs without top-down attention) (detailed in Chapter 1, p. 53-55). Since vOICe is similar to vision in so many areas of perceptual processing, is there a component of neural vOICe processing that is perceptual just like vision? This chapter will investigate whether participants can crossmodally activate the visual cortex with vOICe automatically (without attention), in a similar fashion to perceptual visual processing.

Four functional Magnetic Resonance Imaging (fMRI) tasks will test the hypothesis that vOICe can activate visual cortex without attention. The first task is a passive listening task, where participants detect a pause in the vOICe sound that encodes

a white noise image. The second task is an attention distraction task, where participants count backwards in sevens while a vOICE sound is played. The presence or absence of crossmodal plasticity (*i.e.*, visual activation) in each of these tasks will indicate whether vOICE can be processed in visual regions automatically after vOICE training in comparison to before training. We also tested the specificity of this crossmodal plasticity with a passive listening task for two familiar sounds: a beach sound (as an example of natural sound) and a Star Trek sound (as an example of artificial but familiar sound; task 3). If participants activate visual regions in response to familiar sounds when post-training scans are compared to pre-training scans, it will indicate that the crossmodal plasticity to vOICE is general, not specific to only vOICE sounds, and at least partly automatic. Finally, participants performed a visual control of the first task (passive listening to vOICE), while they detected a pause in the presentation of a white noise image (task 4). This is meant as a direct comparison between the visual response to white noise image and the visual response to the vOICE sound of a white noise image.

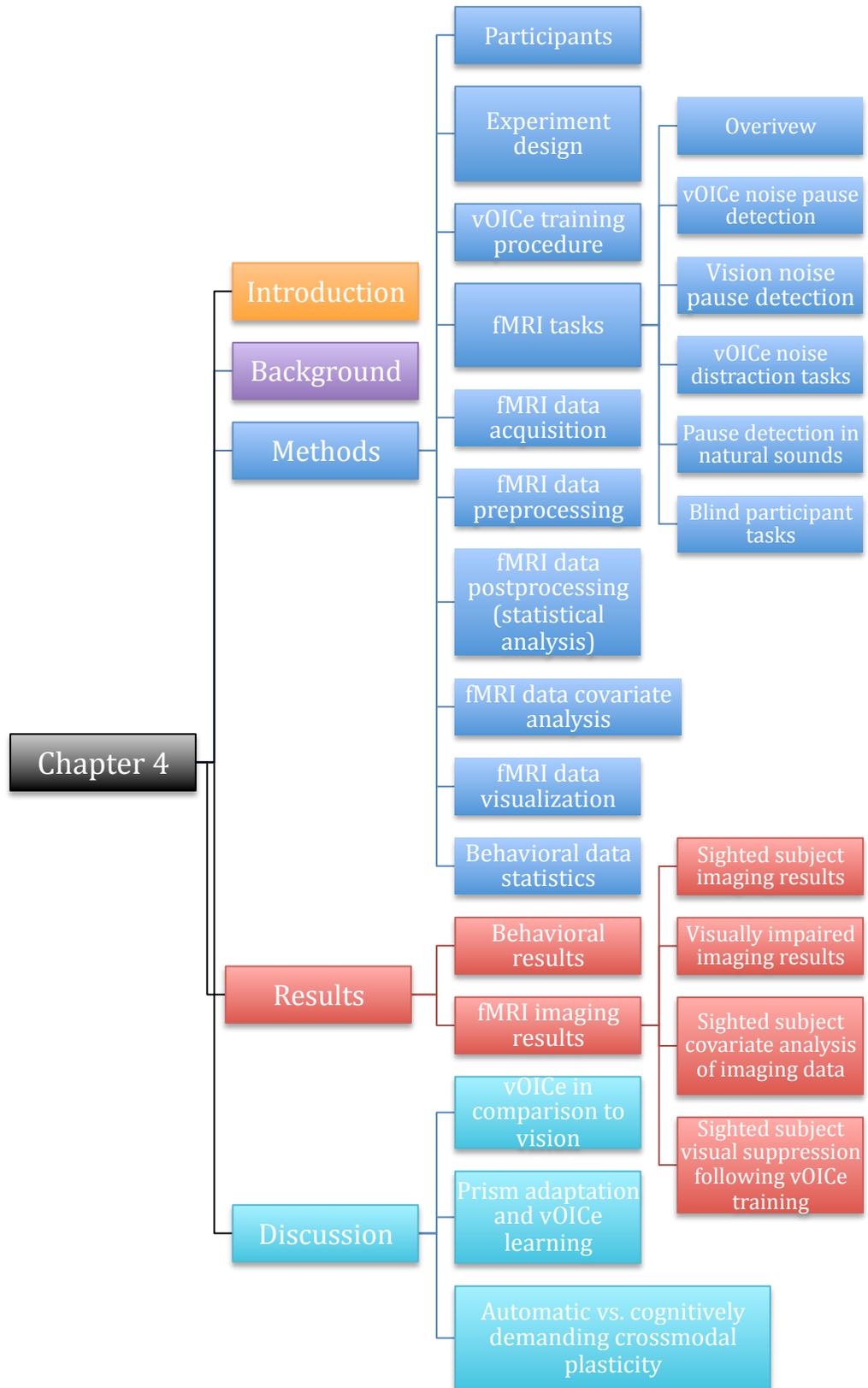


Figure 4.01. Outline of Chapter 4. This figure details the sections of Chapter 4 and their hierarchical structure.

Background

Neural imaging of sensory substitution users has shown crossmodal plasticity in blind and sighted trained users of SS, as discussed in Chapter 1. Studies using functional Magnetic Resonance Imaging (fMRI) have shown visual activation in response to sensory substitution (auditory or tactile) stimuli following training on these devices in blindfolded sighted and blind participants. In particular, imaging studies using pattern recognition and localization tasks with auditory sensory substitution have shown activation in early visual regions such as Brodmann Area (BA) 17 (so-called V1, or the primary visual cortex), BA 18, and BA 19 in blind and blindfolded sighted users (Poirier, De Volder, & Scheiber, 2007).

Surprisingly, despite fairly short training on sensory substitution devices (about 1 to 5 hours) for imaging studies, the crossmodal activation in visual regions seems to be robust. Experiments by Amedi and colleagues and Plaza and collaborators, have shown a functional or task-related activation of visual regions (Amedi, et al., 2007; Plaza, et al., 2009). Further, repetitive Transcranial Magnetic Stimulation (rTMS) studies have indicated that there is a causal relationship between sensory substitution performance and neural activation in visual regions for late and early blind (Collignon, et al., 2007; Merabet, et al., 2009). This may possibly indicate the “metamodal” nature of the visual cortical processing (Pascual-Leone & Hamilton, 2001) (see Chapter 1). Additional background detail on sensory substitution imaging studies can be found in Chapter 1.

Behavioral automaticity studies have been discussed in detail in Chapter 1 and Chapter 3. In general, there are several different criteria for automaticity, including:

“Goal independence criterion,” “the non-conscious criterion,” “speed criterion,” and attentional load sensitivity (Spence & Deroy, 2013). These criteria have been tested with a range of techniques on several different elements of perceptual processing (see Chapter 3 p. 91-92). Many types of visual tasks such as visual search and face perception have been shown to meet some automaticity requirements (see Chapter 1 p. 53-55). As was highlighted in Chapter 3 (p. 91-92), distraction tasks have been used for many multisensory behavioral tests of automaticity (*i.e.*, testing attentional load sensitivity). In particular, in Chapter 3 (p. 95-96), a distraction paradigm was discussed to test the attentional load criteria of automaticity for interpreting vOICe sounds. In the task, participants counted backwards in sevens or performed a visual search task while the vOICe sound was played, and then matched the vOICe sound to one of three images displayed (Chapter 3). It was found that the distraction tasks did not significantly diminish the participants’ ability to match vOICe sounds with images. Therefore, the Chapter 3 experiment indicated that vOICe interpretation can be automatic.

Automaticity studies in the literature have also investigated whether visual neural processing and thereby visual neural activation is independent of attention. As detailed in Chapter 1 (p. 54-55), ignored visual stimuli activation intensity is modulated by attentional load to an alternative task. However, the neural processing of ignored visual stimuli is not eliminated by high attentional load to another visual task. Therefore, the processing of visual stimuli is automatic, but the intensity of that processing may vary with attention. A similar distraction paradigm will be used in this Chapter to test whether attentional load reduces or eliminates the processing of vOICe sounds in visual brain regions.

One challenge to fMRI investigations with sensory substitution is the visualization, or the mental-visual imagery of stimuli by the sighted users of vOICE. Visualization occurs when a short-term memory is spatially re-imagined without direct visual input (Kosslyn & Thompson, 2003). Visual imagery can activate visual cortex, and that activation is retinotopic (activating neighboring neural regions for adjacent regions of visual space) (Klein et al., 2004; Slotnick, Thompson, & Kosslyn, 2005). In a meta-analysis of visualization literature, Kosslyn and Thompson found that several individual conditions significantly correlated with early visual activation during visualization in PET, single photon emission computed tomography (SPECT) and fMRI studies. A few conditions listed by Kosslyn and Thompson included a main task identifying high-resolution image details, a baseline task that is not resting state, a main task testing visual shape properties (not spatial visual properties), a main task lasting 5 minutes or less, or a sensitive neural scanning technique (*i.e.*, 3T or 4T fMRI) (Kosslyn & Thompson, 2003). Visual activation has been shown to occur during a rest or baseline task (with eyes closed) (Kosslyn, Thompson, Klm, & Alpert, 1995). This is likely the reason that studies with a resting baseline and a visualization main task result in visual activation in higher-level visual cortices but not in V1 or V2 (*i.e.*, early visual activation is subtracted out by the baseline). It has also been found that visual-imagery neural activation from short-term memory is stronger than imagery activation from long-term memory (Kosslyn & Thompson, 2003). Interestingly, as mentioned above, visual spatial reasoning, despite often using visualization, does not activate visual cortex, especially when it is not shape-based (Kosslyn & Thompson, 2003).

Visual activation due to visualization has been shown to be causally linked to the

visualization task performance. Kosslyn *et al.* used repetitive Transcranial Magnetic Stimulation (rTMS) to deactivate the calcarine cortex (BA 17) prior to performing a visualization task (Kosslyn et al., 1999). Both the performance of the visualization task was reduced and the performance time (the time to task completion) was prolonged when rTMS was applied to BA 17 in comparison to directed away from BA 17. Further, Farah, Soso and Dasheiff showed in a case study that the “visual angle of the minds eye” was reduced in half horizontally but not vertically when one hemisphere of the occipital lobe was surgically removed from a patient (Farah, Soso, & Dasheiff, 1992). This reduction in size is expected based on the topographical mapping of the left visual field to the right visual cortex, and vice versa. In contrast, there are studies that indicate patients with widespread early visual region damage can often still visualize images. These studies may be an indication of long-term, functional re-organization within the damaged brain, and therefore not negate the necessity of early visual activation for visualization within the normal brain (Kosslyn & Thompson, 2003).

Studies of mental imagery in the blind (late and early onset) have indicated that they can produce, integrate, and manipulate mental images amalgamated from past experience and remaining sensory experience (Cattaneo & Vecchi, 2011). The early blind participant studies have focused on the imagery of tactile shapes and objects. Tactile imagery activated the occipital cortex in early blind participants for “visualization,” or perceptual imagery from shape rotation, tactile texture, and auditory stimuli (Kaski, 2002; Lambert, Sampaio, Mauss, & Scheiber, 2004; Rosler, Roder, Heil, & Hennighausen, 1993; Uhl et al., 1994). In Lambert *et al.*'s study, the name of an animal was listed, and the early blind participants were asked to create a mental image of

that animal. The early blind fMRI data contained neural activation in response to animal imagery including BA 17, BA 18, and BA 19 (with a region of interested analysis). Overall, it seems that the early and late blind can have perceptual imagery that is based on tactile perception and long term memory. In contrast, visualization by sighted individuals relies more heavily on visual spatial information.

The importance of visualization to visual activation by sensory substitution is still under active debate. Poirier *et al.* (2007) argues that visualization is the main method of SS visual activation in sighted sensory substitution users, and crossmodal plasticity is the main method for early blind users (Poirier, De Volder, & Scheiber, 2007). In contrast, fMRI studies using sensory substitution argue that the early blind participants have quite similar imaging results to the sighted participants, and therefore likely used a similar crossmodal method. Since the early blind can't visualize in the same way that the sighted can, never having vision, it not likely that image visualization played an important role in the visual activation from sensory substitution (Amedi, et al., 2007). Other methods used for controlling for visual imagery include additional control participants (not trained on sensory substitution) and auditory (non-sensory-substitution) tasks (Amedi, et al., 2007). Occasionally, a separate visual imagery control task relevant to the main experiment is also used (Merabet, et al., 2009).

Since our fMRI experiments were designed to identify early visual-cortical activation due to vOICe training, a similar possible activation pattern due to visualization is of critical concern. Thus, we designed fMRI scanning experiments with (a) a distraction task, (b) white noise stimuli, (c) early blind participants, and (d) a post-experiment questionnaire on visualization, to address this difficult visualization issue.

We describe the visualization issue as “difficult” because it is very challenging to systematically avoid visualization. Worse than that, more efficient visualization via SS training may not be just an artifact, but rather an intriguing element of the underlying neural mechanisms generating the performance improvement. Utilizing these manipulations (a)-(d) above, we may obtain some indications as to how visualization or other strategies are employed in sighted and blind participants similarly or differently.

Methods

Participants

Ten sighted participants were recruited from the Caltech community (2 Female and 8 Male). All fMRI and behavioral experiments were approved by the Caltech Internal Review Board. All participants had not been trained previously on a sensory substitution device.

One severe low-vision participant (visual acuity: 20/420, Male) and 3 blind participants were recruited from the local blind community (1 Female, 2 Male). The blind individuals were two congenitally blind (WB and SB: Retinopathy of Prematurity, entirely blind since infant) and 1 late blind (Retinitis Pigmentosa, light-perception, 30 years of blindness). The late blind participant had a hearing impairment, and wore a hearing aid. The hearing aid was used during vOICE training, but removed during the fMRI scans; the audio volume was increased during fMRI scanning to compensate. The second congenitally blind (SB) also had a minor hearing impairment in one ear, but did not require a hearing aid. All fMRI and behavioral experiments were approved by the Caltech Internal Review Board for blind participation. The visually-impaired participants had not been trained previously on a sensory substitution device.

Experiment Design

This fMRI experiment has a scan session before vOICe training, followed by vOICe training, and then a scan session following vOICe training, all occurring within two weeks (Figure 4.02). The two scan sessions contain the same tasks in order to capture the participant's neural processing difference due to the training between the scan sessions. The vOICe training lasts for four consecutive days (about one hour per day), and in addition, a short vOICe training session occurs directly before the final scan session, which lasts only 30 minutes. The fMRI scan sessions last two hours each, including experiment setup and audio testing. One participant performed all tasks before and after vOICe training (all comparisons are within subject comparisons).

vOICe Training Procedure

Participants used a vOICe device during auditory sensory substitution training. A detailed description of the vOICe device and general procedures is listed in Chapter 2's methods (p. 63-64).

vOICe device training lasted for about five hours between the pre-training and post-training fMRI scanning sessions. Training was performed for about an hour per day for four days, and a final session on the fifth day of about 30 minutes. Training was performed at a black-felt-covered table, or at a black-felt-covered wall (Figure 4.03 shows the black-felt-covered table). Training sessions began each day with a localization evaluation task, and then continued with localization and recognition training exercises.

The localization task assessed the participant's progress in learning the vOICe translation algorithm and their ability to localize objects with the vOICe device (Figure

4.03 shows task setup). The localization task was performed at the black-felt-covered table. The trainer would place a white circle in one of five locations on a black felt board, and the participant would locate the circle with vOICe, center the circle in the field of view, and then reach for the circle with one finger. The distance between the participant's reach and the circle's center would be measured as a metric of inaccuracy. Feedback was provided to participants by moving their finger from the reached position to the center of the white circle. Thus, the correct direction and location of the circle was provided through tactile and proprioceptive feedback.

The training tasks following the localization task varied from day to day, and progressed from simple to complex (Figure 4.04 for overview and Appendix B for detailed day-by-day tasks). Participants performed both localization and recognition tasks, and transitioned from non-cluttered environments (black felt board) to more cluttered environments (black felt wall: Debris, such as a desk, doorway, and various equipment, was present on the left and right as participants approached the wall. Participants were warned when they focused on debris that the target object was not in view). Training was dynamically adapted to fit the participants learning rate and vOICe interpretation weaknesses. Additional time was spent on tasks of particular difficulty to each participant.

Training attempted to integrate and utilize as many modalities as possible. In the last session of training each day, sighted participants performed on the computer a left-right circle localization task, which asked the participants if a circle is on the left or right in an image or on the left or right in a vOICe sound of that image. The task first displayed just images for the localization task, then played vOICe sounds at the same

time as images presented, and finally just played the vOICE sounds. Although this computer task was relatively simple (just indicating if a circle is on the left or right with vision and/or vOICE), it allows for the integration of information across modalities, which may aid in the development of crossmodal plasticity. (Note: blind participants could not perform this task due to the visual element of the task.)

fMRI Tasks

Overview

Six separate tasks were performed by sighted subjects in each fMRI scanning session. The 4 relevant tasks to Chapter 4 will be described here; the remaining 2 tasks will be detailed in Chapter 5. The blind participants performed 4 separate tasks, 3 of which will be describe in Chapter 4, and 1 of which will be described in Chapter 5.

vOICE Noise Pause Detection

The first task was detection of a pause within a vOICE sound encoding an image of white noise (Figure 4.05). During this task participants fixated on a cross, and listened to a vOICE sound played twice (2 second duration). The vOICE sound's pause was either at the beginning, middle, or end. If the sound had a pause, the participant pressed 1; if there was no pause, the participant responded by pressing 2 (24 percent of trials had no pause, while 76 percent of trials had a pause). The participant performed 50 trials of the pause detection task, and was not told that the noise sound was from the vOICE. This task will be referred to as the vOICE noise or vOICE noise pause detection task in the rest of this chapter.

The vOICe noise was encoded with the vOICe software from a set of 10 white noise images generated in MATLAB. The function “random” in MATLAB was used to generate random numbers in a uniform distribution between 0 and 256 for each element, and then each element value was rounded to the nearest integer. Each element was used to make a matrix of 650 x 795 elements (or pixels). The matrices were converted to grayscale and saved as bmp files. The bmp files were loaded into a .mat file (which was used in the experiment) as truncated images of a 600 x 795 size in order to match the size of the localization images (localization task is detailed in Chapter 5).

Vision Noise Pause Detection

The second task was pause detection of a white noise visual image presentation (same images used in vOICe noise pause detection task) (Figure 4.06). The pause in the image presentation lasted for 0.19 seconds of 2-second continuous image presentation. The participants pressed 1 for pause, and 2 for no pause; 24 percent of trials had no pause, while 76 percent of trials have image pause. The pause could be present at the very beginning, middle beginning or the middle of the image presentation. This task will be referred to as the vision noise or vision noise pause detection task in the rest of this chapter. Note: the main difference between the vOICe noise pause detection and vision noise pause detection is that one task is auditory (vOICe noise pause detection), and the other task is visual (vision noise pause detection).

vOICe Noise Distraction Task

The third relevant task was a distraction task with vOICe sounds (Figure 4.07). Participants were shown a number between or equal to 100 and 149, and were told to count backward from the number in 7s. While participants were counting, a vOICe

sound encoding a white noise image was played. Participants were instructed to ignore such a sound if it did play (thus, no task with the vOICe sounds). Participants were not asked to press buttons during this task. Participants performed 50 trials of the counting distraction task before and after vOICe training. This task will be referred to as the vOICe noise distract or vOICe distract counting task in the rest of this chapter.

Pause Detection with Familiar Sounds

The fourth relevant task was detection of a pause within familiar sounds (same experiment layout as Figure 4.05, but with familiar sounds). Two familiar sounds were used: A sound of a beach (2.04 second duration), and a sound from Star Trek (1.27 second duration). The aim of this task was to determine whether vOICe training affected the neural processing of unrelated familiar sounds. During this task, participants were asked to fixate on a cross in the center of their field of view. Participants were asked to respond by pressing 1 if there was a pause in the sound played, and pressing 2 if there was no pause in the sound. The pause could be present at the beginning, middle, or end of the sound, and the sound was the same duration with and without the pause. The participants performed 50 trials of pause detection for each sound before and after vOICe training. These tasks will be referred to as the Beach noise pause detection, Star Trek sound pause detection, or familiar sounds pause detection task in the rest of this chapter.

Blind Participant Tasks

Blind participants performed the vOICe pause detection, the vOICe noise distraction and the pause detection with familiar sounds tasks. They performed the same three tasks as the sighted participants, except for the way instructions were given, but for the sake of convenience, we count them as different tasks (altogether, 7 tasks for Chapter

4 and 3 tasks for Chapter 5). Instructions for the tasks were read aloud by the Macintosh Computer Speech utility and recorded by QuickTime into an audio mov file. These mov files were converted into wav files, and loaded into MATLAB to be played at the beginning of the experiment. The counting starting numbers for the vOICe distraction task were recorded, saved, and loaded into MATLAB in the same manner, and then programmed to automatically be read aloud at the beginning of each trial. All other elements of the experimental design were the same for the blind participants, including the vOICe training.

fMRI Data Acquisition

A Siemens TIM-Trio 3 Tesla MR scanner in the Caltech Brain Imaging Center (CBIC) was used to collect the neural imaging data. A 12-channel phased-array headcoil and MR Confon headphones were used for data collection and audio delivery, respectively. The imaging parameters were: TR = 2.25 seconds, 38 slices in ascending order, and [3,3,3] millimeter voxel size. Participant responses were recorded with a four-button response box within the scanner, of which two buttons were used. Images were presented with a projector image reflected off of a mirror attached to the headcoil and into the participants view. Eye positioning information was recorded for select tasks using a Restek eye-tracking camera attached to the headcoil, and recorded on a lab computer using PowerDirector software. T1 structural scans were acquired in addition to fMRI functional scans for each participant in either the first or second fMRI scanning session, and were coregistered with functional data. The T1 imaging parameters were: TR = 1.5 seconds, and [1,1,1] millimeter voxel size.

vOICe Experiment Layout

Week 1	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday Pre-training fMRI scan 	Saturday
Week 2	Sunday	Monday vOICe Training (1 hour) 	Tuesday vOICe Training (1 hour) 	Wednesday vOICe Training (1 hour) 	Thursday vOICe Training (1.5 hour) 	Friday vOICe Train. (30 min) Post-training fMRI scan 	Saturday

Figure 4.02. vOICe experiment layout. Schematic diagram showing a typical schedule of the fMRI and vOICe training design. A fMRI scan preceded and followed a training period of four days and about five hours. The fMRI scanning sessions both contained the same tasks, and the vOICe training changed in each session.



Figure 4.03. vOICE localization task setup. Participants performed a localization task to assess their progress on each day of vOICE training. This image depicts the localization task setup, with the vOICE glasses and computer on the table, and the white dot that participants located and reached for on the black-felt-covered wall. The white markers indicate the other four locations at which the white dot can be placed (the markers are not present during the experiment, but rather replaced with nearly invisible black velcro).

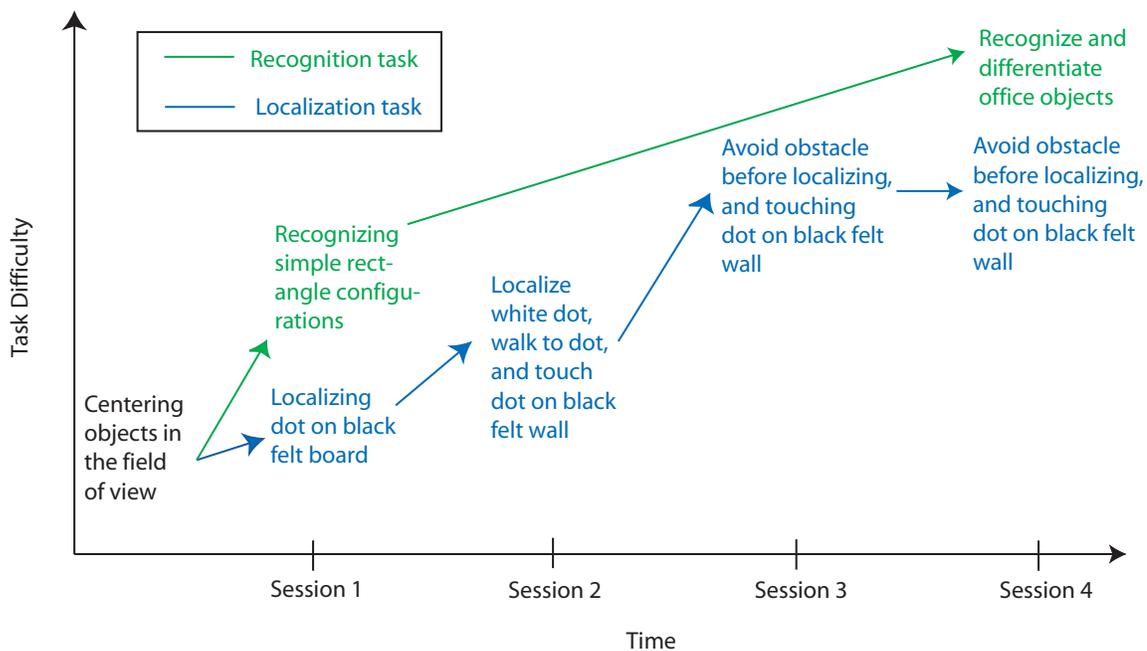


Figure 4.04. vOICE training flow chart. Training was performed on the sensory substitution device (the vOICE) between the pre-training fMRI scan and post-training fMRI scan. This diagram outlines the tasks performed in training and a general time progression of those tasks as a function of difficulty. The localization and recognition tasks are separated into blue and green colors. Each training session was about an hour in duration, but varied to some degree based on each participant's speed at completing each training task.

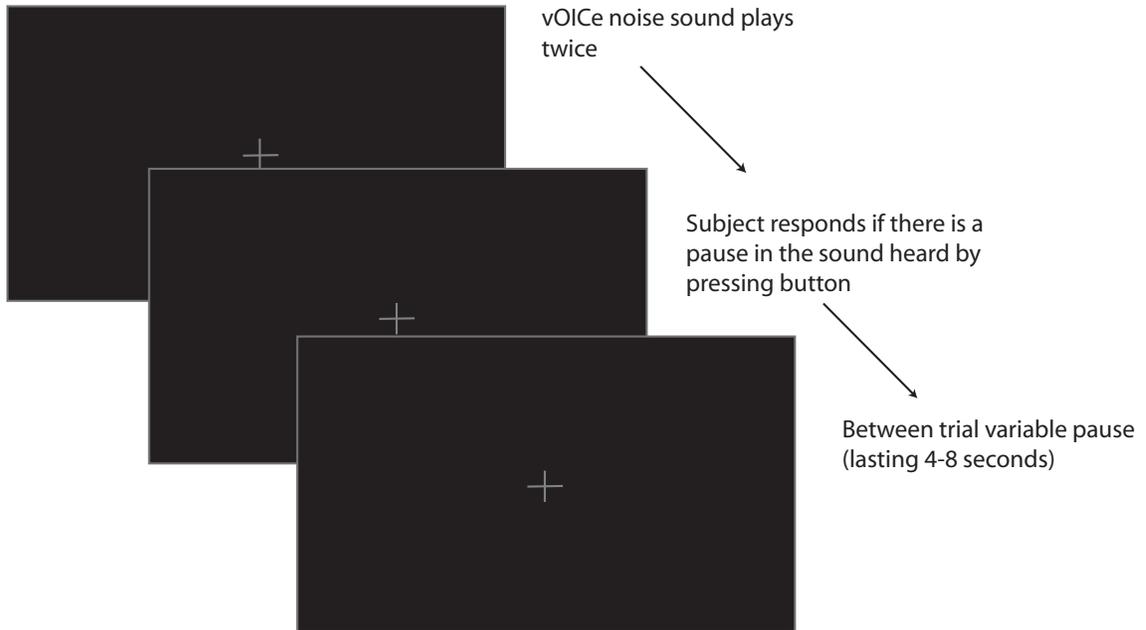


Figure 4.05. fMRI experiment diagram of the auditory pause detection task. Experiment layout of the vOICe noise pause detection task. The pause can be present at the beginning, middle, or end of the vOICe sound. The vOICe sound with the pause is played twice to lengthen the stimulus duration to 2 seconds. The familiar sound pause detection task is designed be the same format as vOICe noise pause detection task. The auditory pause detection task (vOICe noise and the familiar sound pause detection) was performed in both scan sessions.

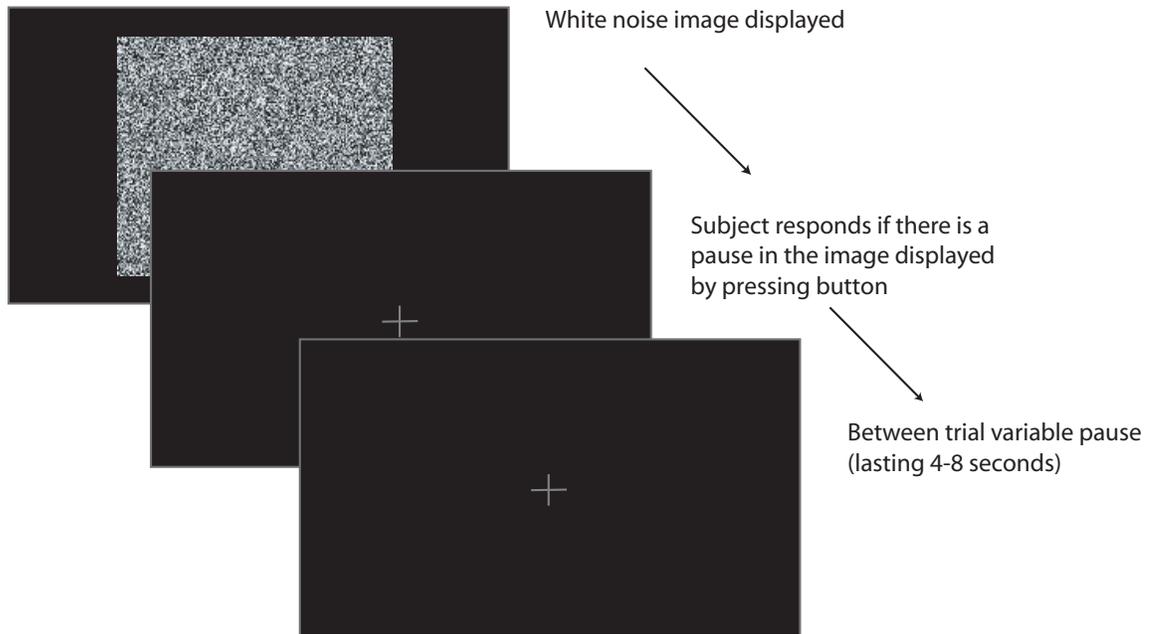


Figure 4.06. fMRI experiment diagram of visual pause detection task. The visual pause detection task used white noise images and asked participants to determine whether the image paused during its presentation. The visual pause detection was used as a control for the vOICE pause detection experiment. The visual pause detection task was performed in both scan sessions.

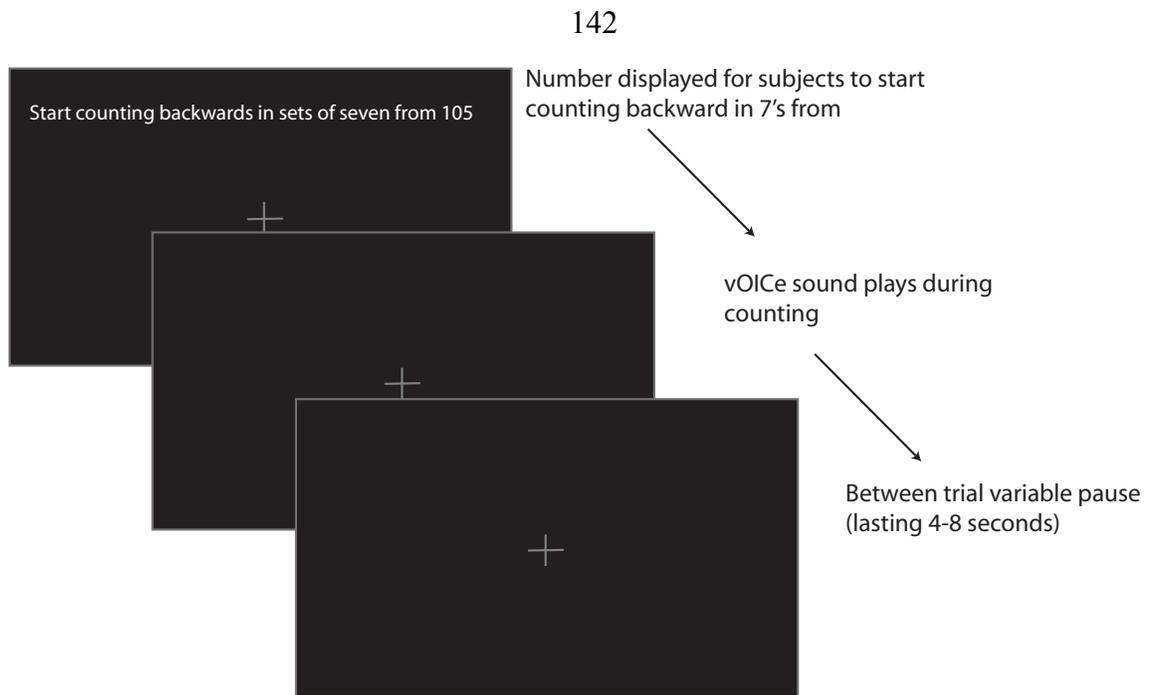


Figure 4.07. fMRI experiment diagram of the vOICE distract counting task. The vOICE distraction counting task presented a number between 100 and 149, and required participants to count backwards in 7s. While participants were counting, a vOICE sound encoding a white noise image was played. Participants were told to ignore the sound played. The vOICE distraction counting task was performed in both scan sessions.

fMRI Data Preprocessing

fMRI preprocessing of the imaging data was performed in SPM8 (The Wellcome Trust Center for Neuroimaging, at the Institute of Neurology at University College London (UCL), UK) (Ashburner et al., 2011). Functional scans were corrected for slice time acquisition, and movement (via image realignment). The co-registration of participant functional and structural images was performed along with normalization to the standard space defined by the ICBM, NIH P-20 project (Ashburner, et al., 2011) and smoothing by an Gaussian kernel of 8 mm full-width at half maximum (FWHM).

fMRI Data Postprocessing (Statistical Analysis)

One general linear model (GLM) was generated, including all 6 fMRI tasks and both the pre-scanning and post-scanning sessions for each participant. High pass filtering was performed in the model specification stage of processing (128 second filter width). No within-participant regressors were used except for standard movement regressors and 12 session constants (one constant per task, 6 tasks pre-training and 6 tasks post-training for sighted participants). The GLM was estimated with a classical algorithm (Restricted Maximum Likelihood). Forty-three contrasts were generated for each sighted participant in order to explore both differences between pre- and post-training conditions, as well as within session comparisons such as localization of a dot on the left vs. the right. The resulting contrasts were summed in a level 2 processing, across all 10 sighted participants. Blind participants were processed individually with a total of 10 contrasts for four tasks in each scan session.

fMRI Data Covariate Analysis

Covariate analyses were used to determine whether any neural activation correlated with a behavioral measurement. A second-level analysis in SPM8 was used in which the contrasts from each participant were summed and a covariate numeric value was entered for each participant in a corresponding matrix. The resulting neural activation from the analysis correlated in strength with the numeric magnitude of the covariate values entered. Covariate values were determined by either the experiment questionnaire (Appendix C) or localization task performance data.

fMRI Data Visualization

Data visualization for both Chapter 4 and Chapter 5 was performed in SPM8. For the section view of Figures 4.12, 4.13, 4.15, 5.5, and 5.7, the neural activation (BOLD functional imaging data) was overlaid on the SPM8 canonical individual T1 structural image. Inflated brain images of Figures 4.12, 4.13, 4.15, 5.5, 5.7, and 5.9 were generated using the render function in SPM8 with the canonical cortical surface image.

Behavioral Data Statistics

ANOCOVA and correlation analyses were performed in MATLAB using the `aoctool`, and `corr` functions.

Results

Behavioral Results

Localization was measured daily during vOICE training at the beginning of each training session (the details of the localization evaluation are in the method section under vOICE training procedure). The inaccuracy of participants' reach for a white circle on a

black felt board with vOICE, normal vision, and random reaching (*i.e.*, no vision or vOICE) is plotted in Figure 4.08 for all of the sighted participants ($N = 10$), and Figure 4.11 for all of the blind participants ($N = 4$). The inaccuracy of the sighted participants' reach decreased with training time (or training sessions, about 1 hour per session) at a rate greater than the random reaching. The slope of the random reaching (*i.e.*, no vOICE or visual input) for the sighted participants is not significantly different from the slope of their vOICE reaching; however, the intercepts are significantly different between random reaching and vOICE reaching (ANOCOVA analysis, $p_{slope} < 0.39$, $p_{Intercept} = 0$). In other words, the sighted participants performed significantly better than random reaching with vOICE at the beginning of training (represented by intercept). However, their learning improvement (represented by slope), while improving at a rate greater than random reaching, was not significantly better. In part, this result is due to the intuitive nature of the localization task; therefore, the task can be learned well in the first half of the first training session trials, generating a large difference between vOICE performance and random reaching. However, as training progresses, participants learn the environment and the most likely spatial regions for the dot location, allowing for improvement at the random reaching control task. This control improvement is then compared to the vOICE task improvement, making it more difficult for the vOICE improvement to be significantly larger. Further, a similar task in the literature by Auvray *et al.* in 2007 showed that reaching for a 4 cm. ball using the vOICE device on a table did not significantly improve in accuracy over two 1-hour training sessions.

Localization improvement with training indicates increased hand-camera (*i.e.*, hand-head) coordination, spatial perception with vOICE, and centering technique (as

described below). When participants begin using the vOICe device (Figure 1.4), they must integrate cognitive information (such as the vOICe encoding principles, and camera location) with perceptual experience and motor commands. Critical elements of that learning process are learning search strategy (the field of view is more limited with vOICe than natural vision), the limits of the camera field of view, the camera position relative to their hand and body, and the sound of different spatial positions such as the top of the field of view, and center of the field of view. The relation between the spatial position of the target in the field of view of vOICe and the field of view in real space can then be used to modulate and guide hand movement during localization.

A training technique of centering then reaching often aids participants in learning vOICe localization. The participant is taught to first locate the object and then center the object in the field of view, therefore identifying the objects position in vOICe coordinates. The participant then tactilely locates the camera on the glasses with their reaching-hand, identifying the direction of their gaze and the physical real-world coordinates of the vOICe field of view. Finally, the participant reaches in the direction that the camera is pointed. This method helps participants improve their accuracy, because it forces participants to consciously note the direction of the camera. Without this conscious reminder of camera direction, participants can easily forget that their head is slightly tilted, altering the location of the center of the field of view in real space. Further, the centering of the sound then camera position identification joins the virtual vOICe space with the real space that the camera's field of view covers, enabling better integration of two types of information.

Learning the spatial limits of target placement is especially relevant to searching

strategy as well as improvement at random reaching. Often, at the beginning of training, participants will not explore the full extent of the black felt board where the target is located; rather, they will explore only the upper or lower half, or left or right half of the board. When they begin to become confused or frustrated at not locating the target within their limited search radius, the experimenter often provides a hint of the unsearched section of board. As the participant progresses in the task, they learn where the target can and cannot be located, reducing confusion in their search strategy. This learning of the limits of the potential target locations is likely the reason that random reaching improves slightly with training time, as participants will likely direct even their random reaching to the general area of target locations. By reducing search time in unnecessary spatial locations, learning the target space is also important to improving participants' task efficiency.

Sighted participant localization task performance also seems to show a ceiling effect. The sighted participant individual localization performance vs. training session is plotted in Figure 4.09. While participants have a wide range of starting localization accuracy in session 1, their final localization performance range has narrowed to a much smaller range. Another way of representing this effect is in Figure 4.10, where the slope and intercept for each participant are plotted on a scatter plot (each data point is a different participant). Interestingly, the slope and intercept across participants are significantly negatively correlated ($\rho = -0.7464, p < 0.01$). This means that when the starting position (*i.e.*, intercept), increases (*i.e.*, becomes more positive, more inaccurate) the slope decreases (*i.e.*, becomes more negative, or higher learning rate). Therefore, independent of where participants begin their performance of the localization task, the

learning rate compensates to make their end performance within a small range of localization accuracy. This effect may be due to a limiting factor that prevents the early-high-performance participants from improving the same amount as the early-low-performance participants. It may be true that the participants performing better at the beginning of training better translate their cognitive knowledge to spatial interpretation and hand-camera coordination in contrast to those that perform worse at the beginning of training. However, as training continues, both the early and late learners are limited by the resolution of vOICE vertically and horizontally, hand-camera coordination, and the lack of visual feedback during the reaching movement. It is useful to note, though, that this “ceiling” could still possibly be overcome by longer, more intensive training than used in this experiment.

Low-vision and blind participants were also trained on the vOICe device, and performed the localization task (Figure 4.11). Panel A of Figure 4.11 shows the localization results for a severe low-vision participant, FZ (visual acuity: 20/420), with late-onset visual impairment. His performance indicates a rapid rate of learning (FZ slope: -1.36 , the more negative the better), much greater than the sighted participants (sighted participant slope: -0.63); however, his initial performance is also much worse (FZ intercept: 11.22 inches, sighted participant intercept: 7.38 inches). His results thus follow a qualitatively similar trend of the “ceiling effect” to that in the sighted. The blind participants ($N = 3$) (participant details in methods) performed similarly to the sighted on the improvement of localization with vOICe (blind vOICe slope: -0.41), but had a larger slope for the control random reaching than the sighted (blind control slope: -0.55 ; sighted control slope: -0.23). It is unclear exactly why the blind participants random reaching improved so much; it may be an artifact of the limited number of trials (only 10 trials per day). It is also possible that training on the vOICe device improved the blind participants’ deficit in spatial awareness. This improved spatial awareness could have been measured as an improved sense of the target space that they unconsciously reached toward during random reaching. The final possibility is that because the blind have a heightened sense of hearing, they may (toward the end of training) have been able to barely hear the general direction of the dot placement.

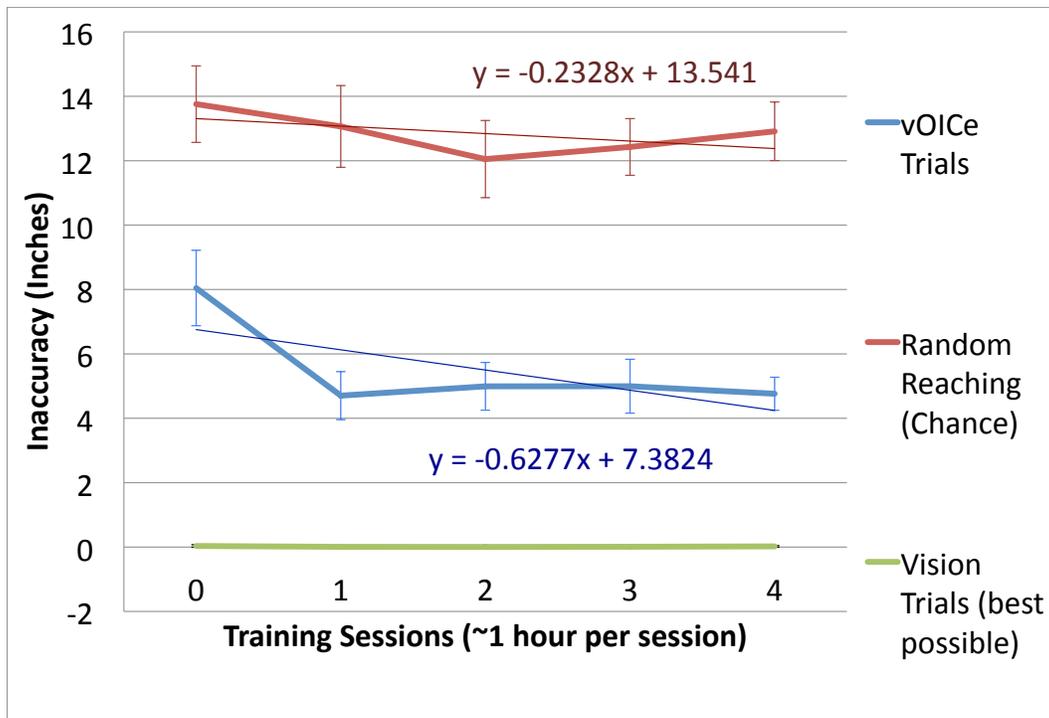


Figure 4.08. Sighted participant localization behavioral performance. Sighted fMRI participants ($N = 10$) performed a localization task in each training session (Figure 4.03), where they reached for a white dot on a black felt board. The spatial inaccuracy of their reach was recorded in inches. This inaccuracy is plotted when using vOICE to localize the dot, when using vision to localize the dot, and when using neither vOICE or vision (*i.e.*, random reaching).

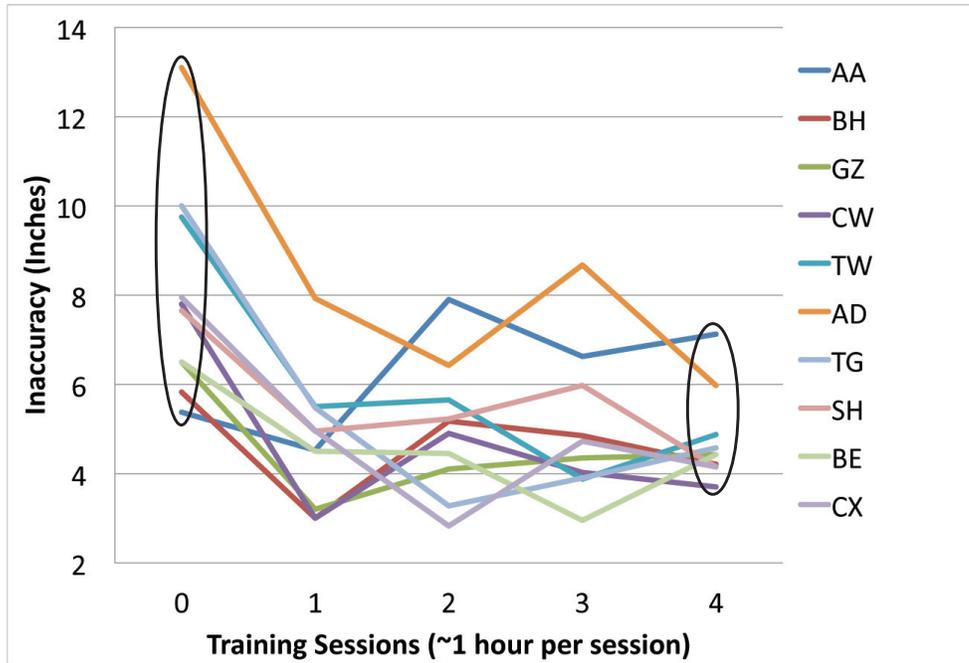


Figure 4.09. vOICE localization behavioral performance in individual participants. Sighted fMRI participants ($N = 10$) performed a localization task in each training session (Figure 4.03), where they reached for a white dot on black felt. The spatial inaccuracy of their reach was recorded in inches. This inaccuracy is plotted when using vOICE to localize the dot, when using vision to localize the dot, and when using neither vOICE or vision (*i.e.*, random reaching) in Figure 4.08. This plot shows the individual participants' performance at the vOICE device alone. This plot shows that the range of initial localization inaccuracies (ellipse on left) is much wider than the range of final inaccuracies (ellipse on right). The narrowing in performance range with training sessions supports the ceiling-effect hypothesis in Figure 4.10.

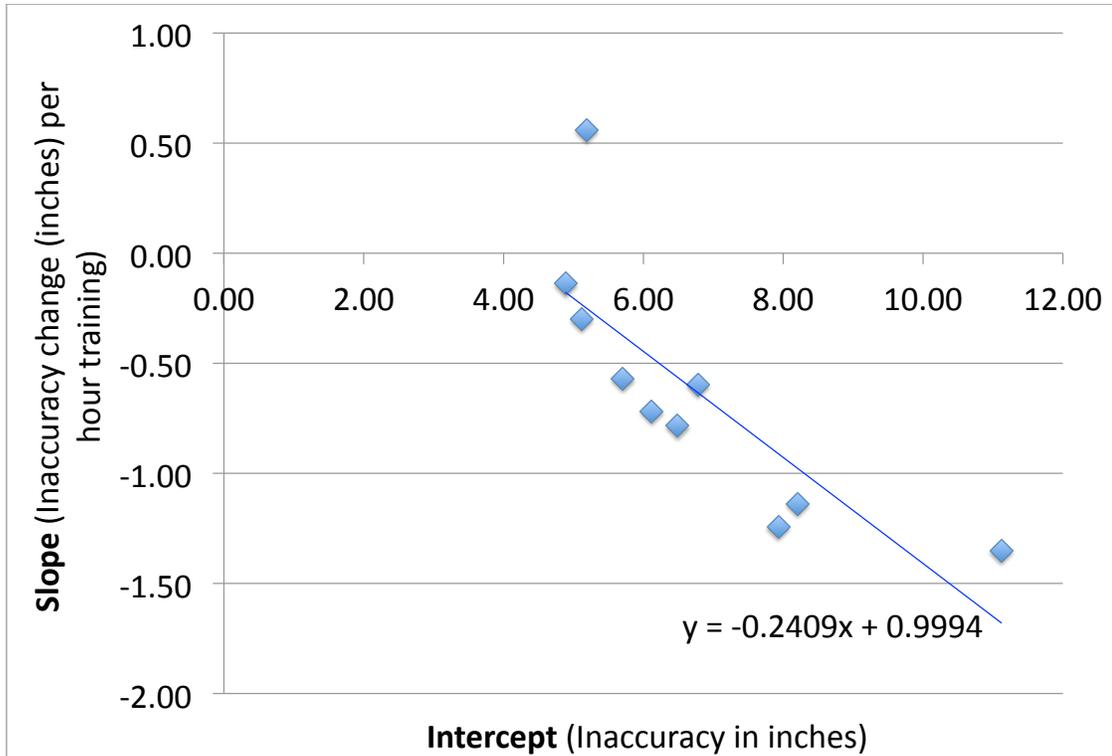
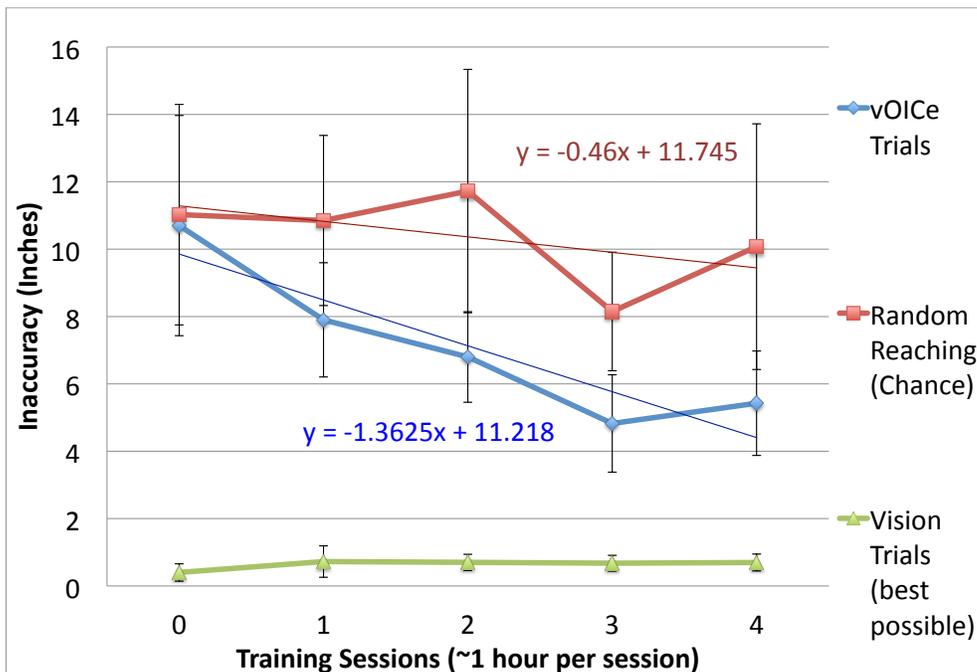


Figure 4.10. Sighted localization behavioral performance, slope vs. intercept. fMRI participants performed ($N = 10$) a localization task in each training session (Figure 4.03), where they reached for a white dot on black felt. The spatial inaccuracy of their reach was recorded in inches. The slope and intercept of this vOICe localization inaccuracy vs. training session plot for the individual participants is plotted above (each data point is a different participant) ($\rho = -0.7464$, $p < 0.01$). The correlation between the slope and intercept of participants' localization performance indicates a possible ceiling effect, where the performance of participants with initially low inaccuracy improved at a slower rate (less negative slope) than participants with initially high inaccuracy. In effect, all participants asymptoted to a similar final performance, independent of their initial inaccuracy with vOICe localization. This indicates that specific elements of this vOICe training or the vOICe in general are limiting further improvement. These limitations

could possibly be overcome with more extensive training.

A



B

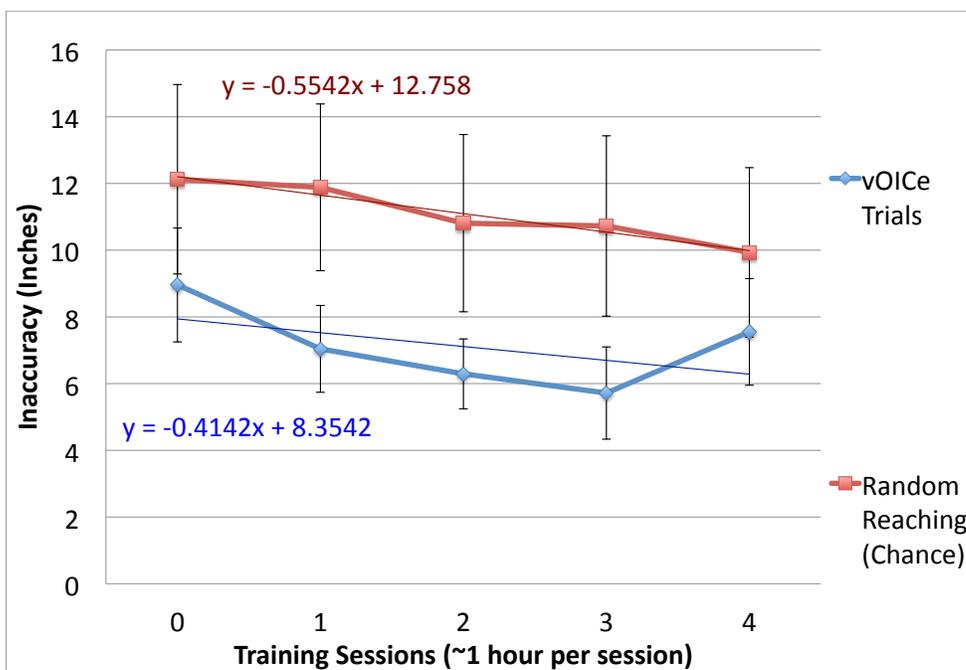


Figure 4.11. Blind and severe low-vision participant localization behavioral performance. Panel A shows a severe low-vision fMRI participant ($N = 1$) localization task performance. Panel B shows blind fMRI participants ($N = 3$; one late blind and two congenitally blind) localization task performance. In each session (Figure 4.03), participants reached for a white dot on black felt. The spatial inaccuracy of their reach was recorded in inches. This inaccuracy is plotted when using vOICe to localize the dot, when using neither vOICe or vision (*i.e.*, blind trials or random reaching), and when using vision alone (only with the low-vision participant).

fMRI Imaging Results

Sighted Participant Imaging Results

Two fMRI tasks are the primary focus of testing the automaticity of crossmodal plasticity with vOICE: The vOICE-noise pause-detection task, and the vOICE-noise distraction task (see the methods for the task details). The contrast of vOICE noise pause detection [Post-training – Pre-training] was used to test whether training on vOICE induced crossmodal plasticity with visual activation (Note: [Post-training – Pre-training], and [Post – Pre] will both be used to indicate that the pre-vOICE-training scans were subtracted from the post-vOICE-training scans). Figure 4.12 and Table 4.1 show the results for this contrast on 10 sighted participants. Sighted participants were found to have significant activation in Brodmann Area (BA) 39 for this task. A small volume correction for BA 39 yielded pvalues less than 0.05, when a sphere of 10 millimeter radius (mm) was used (Table 4.1).

The contrast of vOICE noise distraction task [Post-training – Pre-training] was used to investigate the impact of vOICE training on automatic crossmodal visual activation in 10 sighted participants. Figure 4.13 and Table 4.2 show the results for this contrast. Significant activation was found in Brodmann Areas (BA) 41, 19, and 18 among others. A small volume correction for BA 19 and 18 yielded pvalues not less than 0.05, when a sphere of 10 mm radius was used (Table 4.2). Nonetheless, a small volume correction with 5 mm radius sphere does yield pvalues less than 0.05 (BA 19 [–39 –76 –2] $p < 0.02$, BA 19 [–33 –67 –2], $p < 0.04$, BA 18 [–24 –79 19] $p < 0.03$). Therefore, the small variation in the stringency of the multiple comparisons correction makes the

activity significant, which may indicate relatively smaller cortical volumetric changes as the neural correlates of the training effect.

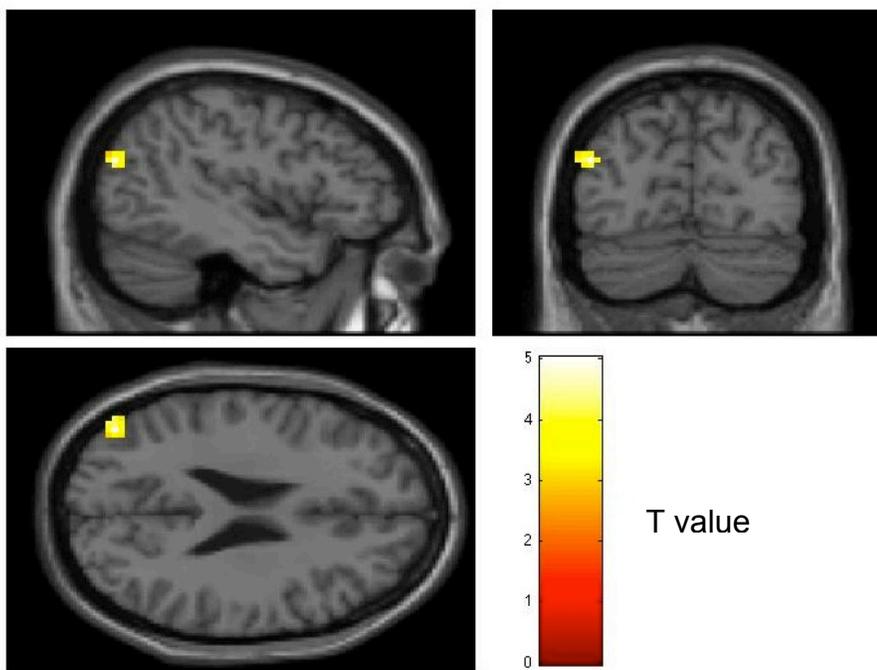
The neural activation results for the pause detection task with familiar sounds are presented in Table 4.3. The contrast of familiar sounds pause detection [Post-training – Pre-training] was used to identify changes in neural processing engendered by the training on the vOICe device. For this Post-Pre contrast, BA 39 and 40 were activated similarly to the vOICe noise pause detection task Post-Pre contrast, but no early visual areas were activated.

In both the vOICe-noise pause detection and the familiar-sound pause-detection tasks, BA 39 was significantly activated in a post-training minus pre-training contrast. BA 39, as a part of the angular gyrus, is known for many different types of functions ranging from language processing, calculation, and visual spatial processing (Bernal & Perdomo; Delazer et al., 2003; Inui et al., 1998; Kohler, Kapur, Moscovitch, Winocur, & Houle, 1995). While the visual spatial processing is most relevant to the task in this experiment, the other functions can not be entirely ruled out, though improbable (the participant was not reading or calculating). In crossmodal interactions and sensory substitution fMRI studies, BA 39 is also a frequent participant. As mentioned earlier, when the angular gyrus is damaged, participants lose the bouba-kiki effect, a common and strong shape-to-sound crossmodal mapping (Ramachandran & Hubbard, 2003). In sensory substitution studies, BA 39 is frequently activated during SS interpretation tasks (Plaza, et al., 2012; Poirier, De Volder, Tranduy, et al., 2007). Therefore, it is likely that BA 39 is mediating the crossmodal integration that is essential to the spatial and visual interpretation of vOICe sounds.

The vOICe distraction task (counting backwards) activated BA 19 and 18 when the pre-training scans were subtracted from the post-training scans. This visual activation in early visual areas (V3 and V2) could be generated by visual imaginings of numbers and shapes. While the post- and pre-scanning session subtraction should remove this visual imagining (there is no reason the visual imagining should not occur in both sessions and therefore cancel out), it is also possible that vOICe training strengthened visualization, making it easier after training. An experiment questionnaire and visualization covariate will further answer these questions in the following pages.

If the visual activation in the vOICe distraction task [Post – Pre] is due to the vOICe training, then the visual activation shows that crossmodal plasticity can be activated automatically (*i.e.*, with attentional distraction). As highlighted in the introduction to the chapter, this is an entirely new result to the sensory substitution field, and indicates that sensory substitution processing might not be entirely processed in a cognitive top-down fashion. It shows that the crossmodal plasticity is resistant to attentional load and therefore at the neural level acts more like vision than was ever suspected. This result may be the critical first step in generating SS training procedures and encoding algorithms that capitalize on this automatic crossmodal processing to obtain stronger, more intuitive SS interpretation and use.

A



B

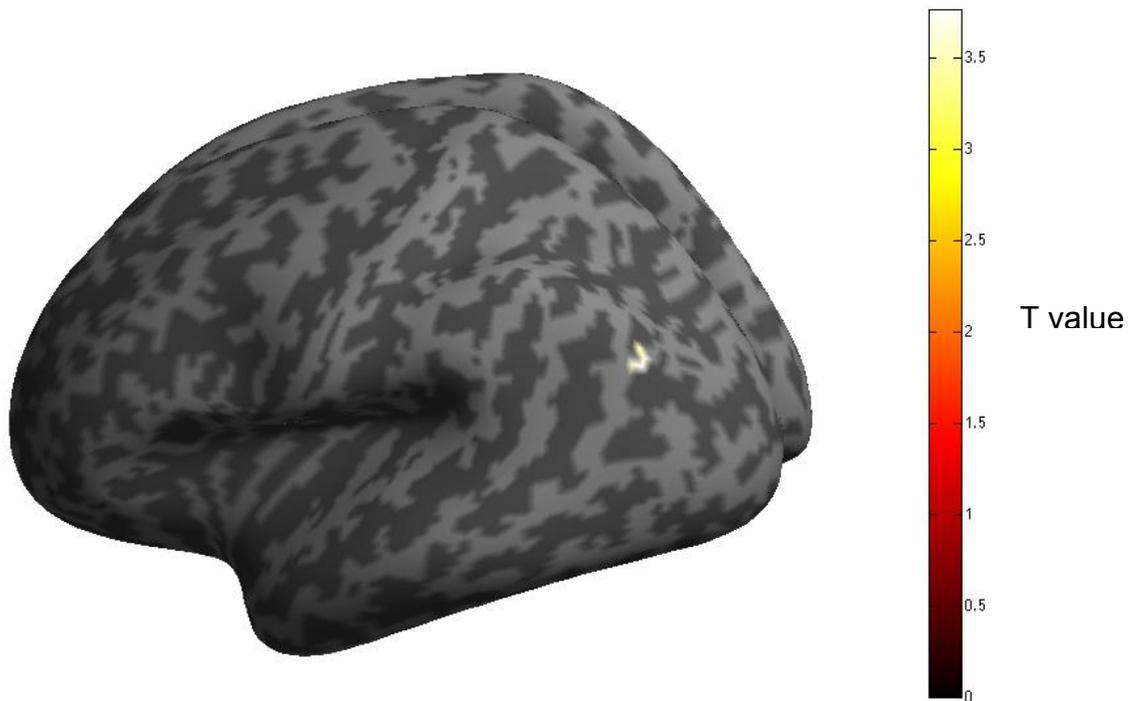
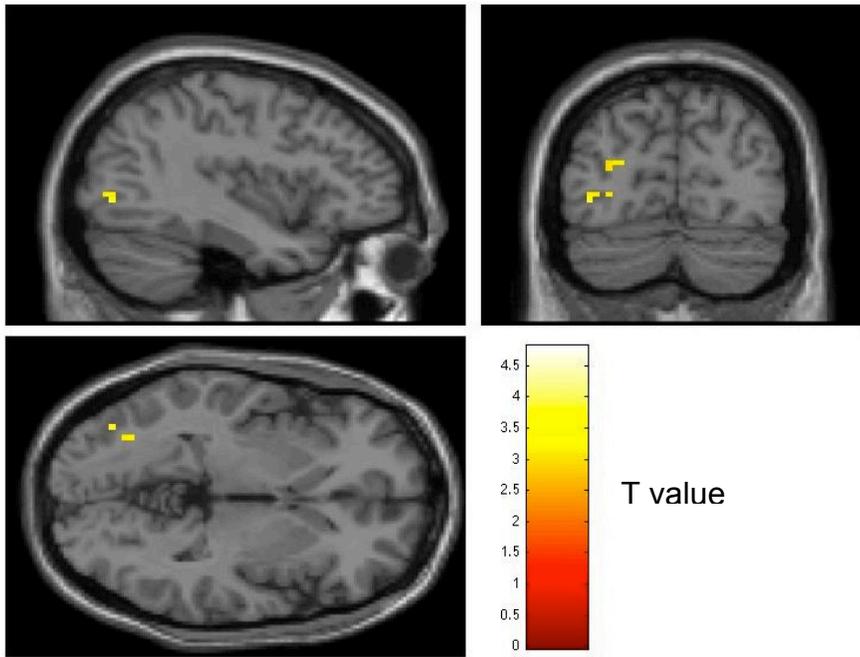


Figure 4.12. fMRI data: Post – pre training vOICE noise sighted participants. The neural imaging result is displayed for post-vOICE-training in contrast to pre-vOICE-training for the vOICE noise pause detection task in sighted participants ($N = 10$). Imaging data presented shows activation in BA 39, and is $p < 0.009$ uncorrected and clusters of 10 voxels or more. Further correction for multiple comparisons is shown in Table 4.1. The detailed description of methods for fMRI data display are in the Chapter 4 methods.

Sighted Participants ($N = 10$)						
Region	BA	Side	x	y	z	p_{uncorr}
vOICe Noise Pause Detection [Post – Pre]						
Middle Temporal Gyrus	39	L	-45	-76	25	0.000
- small volume-corrected peak						0.029*

Table 4.1. fMRI data: Post – pre training vOICe noise pause detection sighted participants. Imaging results for sighted participants when comparing post-vOICe-training scan and the pre-vOICe-training scan ($N = 10$). All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

A



B



Figure 4.13. fMRI data: Post – pre training vOICe-noise distract task in sighted participants. The neural imaging result is displayed for post-vOICe-training in contrast to pre-vOICe-training for the vOICe-noise distract task in sighted participants ($N = 10$). Imaging data presented shows activation in BA 19 and 18 among other regions, and is $p < 0.009$ uncorrected and clusters of 10 voxels or more. Further correction for multiple comparisons is shown in Table 4.2. Methods for fMRI data display are in the Chapter 4 methods.

Sighted Participants ($N = 10$)						
Region	BA	Side	x	y	Z	p_{uncorr}
vOICe Distract Counting [Post – Pre]						
Superior Temporal Gyrus	41	R	39	-31	4	0.000
Inferior Occipital Gyrus	19	L	-39	-76	-2	0.003
<i>- small volume-corrected peak</i>						0.074*
Lingual Gyrus	19	L	-33	-67	-2	0.006
<i>- small volume-corrected peak</i>						0.117*
Cuneus	18	L	-24	-79	19	0.004
<i>- small volume-corrected peak</i>						0.096*
Middle Occipital Gyrus	19	L	-33	-73	19	0.006
Middle Occipital Gyrus	19	L	-33	-79	13	0.008
Posterior Cingulate	29	L	0	-52	10	0.005
Posterior Cingulate	30	L	0	-43	19	0.006

Table 4.2. fMRI data: Post – pre training vOICe noise distract task in sighted participants. Imaging results for sighted participants when comparing post-vOICe-training scan and the pre-vOICe-training scan ($N = 10$). All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within ± 5 mm are not included.

Sighted Participants ($N = 10$)						
Region	BA	Side	x	y	z	p_{uncorr}
Beach Sound Pause Detection [Post – Pre]						
No Activation						
Star Trek Sound Pause Detection [Post – Pre]						
Insula	13	R	39	-46	19	0.000
Middle Temporal Gyrus	39	R	45	-55	7	0.001
- small volume-corrected peak						0.033*
Thalamus		R	6	-28	10	0.000
Caudate		R	21	-40	10	0.000
Thalamus		L	-6	-34	10	0.000
Middle Frontal Gyrus	6	R	33	-1	64	0.000
Caudate		R	3	5	4	0.000
Caudate		R	3	17	7	0.003
Precuneus	7	R	21	-49	46	0.000
Inferior Parietal Lobule	40	R	33	-43	46	0.001
Inferior Parietal Lobule	40	R	39	-55	46	0.004
Precentral Gyrus	6	L	-24	-16	70	0.001
Precentral Gyrus	6	L	-33	-7	67	0.005
Medial Frontal Gyrus	8	L	-12	38	34	0.001
Postcentral Gyrus	5	L	-24	-43	58	0.001

Table 4.3. fMRI data: Post – pre training familiar sounds sighted participants. Select imaging results for sighted participants when comparing post-vOICE-training scan and the pre-vOICE-training scan ($N = 10$) (only the top 15 clusters of activation are presented in this table; a full list is in Appendix D, Table A). All regions were limited to $p < 0.009$

uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

Visually-Impaired Imaging Results

The visually-impaired participants ranged from severe low vision (participant FZ) ($N = 1$) to total blindness ($N = 3$) (participant details and method alterations to accommodate the blind are in the Methods section). These participant groups were tested for their similarity and differences in crossmodal plasticity and neural processing of vOICe in general. Crossmodal plasticity in the blind is discussed in detail in Chapter 1 (p. 48-49), including visual cortical activation during braille reading. It is also mentioned in Chapter 1 (p. 42) that deactivation of visual cortex with repetitive Transcranial Magnetic Stimulation (rTMS) causes a decrease in sensory substitution performance in the blind, but not in the sighted sensory substitution users. This existing literature indicates that neural plasticity and multimodal integration can be quite different in the blind relative to the sighted, and therefore it is important to compare them directly.

The severe low-vision participant, FZ (visual acuity: 20/420), with late onset visual impairment, performed the fMRI experiment; his neural imaging results are presented in Table 4.4, panel A. In the vOICe noise and familiar sound pause detection, and the vOICe distraction task (post training – pre training) participant FZ had neural activation in Brodmann Area 40. BA 40 is a region previously found to process sensory substitution (SS) and to integrate multisensory information. In sensory substitution processing, BA 40, was found in a study by Ortiz *et al.* to be a significant difference between the blind (with no visual “experience” from SS) and the blindfolded sighted following tactile SS training (Ortiz, et al., 2011). Ortiz *et al.*’s result is similar to our result of BA 40 activation in the fMRI imaging of a nearly blind participant during vOICe tasks. BA 40 was also found to be active in a depth perception study using

auditory SS on sighted participants (Renier, Collignon, Poirier, Tranduy, Vanlierde, Bol, Veraart, & De Volder, 2005). BA 40 is “known” as a multisensory region with “super-additive” response to audiovisual speech stimuli (Calvert, Campbell, & Brammer, 2000; James & Stevenson, 2012). Other functions of BA 40 are writing, language comprehension, memory, calculation, motor planning, and music performance (Bernal & Perdomo, 2014). Many of these functions can be ruled out due to the participant, FZ, not reading, listening to language, or performing musically during the tasks in question. Memory of the vOICe training, motor planning, and calculation are possible, but do not apply to all the conditions in which BA 40 was activated (in the distraction task, participants do not press buttons, and are distracted from attentional interpretation of vOICe), whereas multisensory processing does apply to all conditions, making multisensory processes the most probable function of BA 40 in this study.

A late-blind participant, RD, also performed the vOICe fMRI experiment (results in Table 4.4 B; blindness details in methods). Similar to the severe low-vision participant, the late blind participant had activation in BA 40 for all task contrasts, including the vOICe Noise Pause Detection [Post – Pre], the vOICe Distract Counting [Post – Pre], the Beach Sound Pause Detection [Post – Pre], and the Star Trek Sound Pause Detection [Post – Pre]. In addition to this, the late blind participant had activation in BA 39 in the vOICe Noise Pause Detection [Post – Pre], the Beach Sound Pause Detection [Post – Pre], and the Star Trek Sound Pause Detection [Post – Pre]. Brodmann Area 39 is a multimodal region also activated in the sighted participants’ ($N=10$) imaging results (Table 4.1, Figure 4.12), and is discussed in detail on p. 157. Finally, the late blind participant had neural activation of at least one early visual region, (*i.e.*, BA 17,

18 and 19) in each of contrasts of interest (Table 4.4 B). Therefore, the late blind participant appears to have vigorous crossmodal plasticity that activated early visual regions with vOICe and familiar sound stimuli. This vigorous crossmodal plasticity would be expected in a visually deprived individual, especially one that is late blind. In addition, the late blind participant utilized multisensory regions (such as BA 40) for processing the auditory vOICe input; this would be expected in a late blind participant with normal multimodal integration between vision and audition generated before the onset of blindness.

Two congenitally blind participants performed the vOICe neural imaging experiment (results in Table 4.4 C-D; blindness details in methods). The first, WB, had neural activation in BA 19 for vOICe Distract Counting Task with a [Post – Pre] contrast (Table 4.4 C). Therefore, vOICe auditory stimuli automatically activated visual regions in WB, just like automatic crossmodal activation seen in the late blind participant (Table 4.4 B) and the sighted participants (Table 4.2). Activation in BA 19 has been shown in many sensory substitution imaging studies, as described in Poirier *et al.*'s literature review (Poirier, De Volder, & Scheiber, 2007). BA 19 has also been found to be active during braille reading in the blind (Burton et al., 2002). Despite congenitally blind participant WB's visual activation in response to vOICe stimuli (via crossmodal plasticity), he lacked multimodal region activation (such as BA 40 or BA 39) in response to vOICe stimuli. The absence of multimodal region activation may be due to his limited experience with vision and audition interactions. In general, it is likely that a congenitally blind individual has underdeveloped (or absent) multimodal neural processing between these two types (*i.e.* spatial and temporal) of modalities. Therefore,

the absence of multimodal region activation in response to vOICe in a congenitally blind participant (when compared to the sighted and late blind participants) is not surprising.

The second congenitally blind participant, SB, had neural activation in BA 18 for the vOICe Noise Pause Detection Task with a [Post – Pre] contrast, and no visual activation for the vOICe Distract Counting Task also with a [Post – Pre] contrast (Table 4.4 D). Therefore, crossmodal plasticity was less likely to be automatic in participant SB. However, in comparison to participant WB (congenitally blind), SB did have multimodal neural activation in BA 40 for several of the contrasts. In a congenitally blind individual, BA 40 may have been taken over by auditory or tactile processing, and therefore indicate a different type of processing than in a sighted or late blind individual.

Overall, two out of three of the blind participants ($N = 3$) had visual activation in the vOICe distract counting [Post – Pre] contrast, indicating automatic processing of vOICe in visual regions. However, in the severe low-vision participant, the vOICe distract counting [Post – Pre] contrast generated BA 40 activation but no early visual activation, meaning that the processing of SS was performed primarily in a multisensory region, rather than in multisensory and visual regions. This difference may be due to the different neural architecture of the low-vision brain (compared to the blind), though it may also be an individual difference. It cannot be conclusive with only one low-vision participant. In general, it can be concluded that the vOICe is processed automatically (*i.e.*, independent of cognitive load) in either multisensory or visual regions for most of the visually-impaired participants tested ($N = 4$).

A

Severe Low-Vision Participant (N = 1) (FZ)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
vOICe Noise Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	R	48	-37	37	0.000
<i>- small volume-corrected peak</i>						0.000*
Inferior Parietal Lobule	40	R	57	-43	40	0.000
Superior Frontal Gyrus	6	R	9	14	58	0.001
Superior Frontal Gyrus	6	L	-6	8	58	0.002
vOICe Distract Counting [Post – Pre]						
Inferior Parietal Lobule	40	R	57	-43	40	0.000
<i>- small volume-corrected peak</i>						0.000*
Inferior Parietal Lobule	40	R	51	-34	37	0.002
Postcentral Gyrus	3	L	-21	-34	70	0.002
Star Trek Sound Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	R	57	-43	40	0.000
<i>- small volume-corrected peak</i>						0.006*
Inferior Parietal Lobule	40	R	48	-34	37	0.001
Beach Sound Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	R	54	-46	40	0.000
<i>- small volume-corrected peak</i>						0.008*

B

Late Blind Participants (N = 1) (RD)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
vOICe Noise Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	R	69	-25	25	0.000
<i>- small volume-corrected peak</i>						0.000*
Precentral Gyrus	4	R	60	-7	22	0.000
Supramarginal Gyrus	40	R	51	-52	25	0.000
Inferior Parietal Lobule	40	L	-60	-28	28	0.000
Supramarginal Gyrus	40	L	-48	-49	34	0.000
Supramarginal Gyrus	40	L	-42	-37	34	0.000
Middle Temporal Gyrus	39	L	-45	-67	25	0.000
<i>- small volume-corrected peak</i>						0.000*
Caudate		R	21	-1	22	0.000
Caudate		R	18	8	22	0.000
Cingulate Gyrus	24	R	24	-10	34	0.000
Superior Frontal Gyrus	8	R	18	38	52	0.000
Middle Frontal Gyrus	8	R	24	38	40	0.003
Lingual Gyrus	19	R	33	-61	1	0.000
<i>- small volume-corrected peak</i>						0.009*
Caudate		L	-15	8	19	0.000
Caudate		L	-18	-16	22	0.002
vOICe Distract Counting [Post – Pre]						
Middle Temporal Gyrus		R	51	-34	1	0.000
Superior Temporal Gyrus		R	63	-16	-2	0.000
Cuneus	17	R	12	-82	10	0.000
<i>- small volume-corrected peak</i>						0.000*

Late Blind Participants (N = 1) (RD) Continued						
Region	BA	Side	x	y	Z	<i>p</i>_{uncorr}
Posterior Lobe, Cerebellum		R	30	-64	-8	0.000
Posterior Lobe, Cerebellum		R	21	-76	-14	0.000
Insula	13	R	48	-22	25	0.000
Inferior Parietal Lobule	40	R	66	-37	28	0.000
- small volume-corrected peak						0.000*
Inferior Parietal Lobule	40	R	39	-52	43	0.000
Middle Frontal Gyrus	8	L	-33	35	43	0.000
Middle Frontal Gyrus	8	L	-30	26	40	0.000
Middle Frontal Gyrus	9	L	-39	38	34	0.000
Inferior Parietal Lobule	40	L	-54	-28	25	0.000
Insula	13	L	-45	-19	19	0.000
Cingulate Gyrus	32	L	0	17	40	0.000
Medial Frontal Gyrus	6	L	-9	-4	58	0.000
Beach Sound Pause Detection [Post – Pre]						
Precuneus	19	L	-24	-85	43	0.000
- small volume-corrected peak						0.000*
Supramarginal Gyrus	40	L	-60	-46	37	0.000
- small volume-corrected peak						0.000*
Superior Occipital Gyrus	19	L	-36	-82	34	0.000
Middle Temporal Gyrus	39	R	45	-61	28	0.000
- small volume-corrected peak						0.001*
Inferior Parietal Lobule	40	R	69	-25	25	0.000
Precuneus	19	R	33	-79	34	0.000
Middle Frontal Gyrus	8	L	-45	17	49	0.000

Late Blind Participants (N = 1) (RD) Continued						
Region	BA	Side	x	y	Z	<i>p</i>_{uncorr}
Superior Frontal Gyrus	8	L	-27	44	40	0.000
Superior Frontal Gyrus	9	L	-18	59	34	0.005
Superior Frontal Gyrus	9	L	-27	56	34	0.008
Superior Frontal Gyrus	10	L	-42	50	25	0.000
Lingual Gyrus	19	L	-33	-67	-2	0.004
Star Trek Sound Pause Detection [Post – Pre]						
Cuneus	17	R	9	-82	10	0.000
- small volume-corrected peak						0.000*
Lingual Gyrus	18	L	-15	-79	-5	0.000
- small volume-corrected peak						0.003*
Lingual Gyrus	18	R	18	-70	4	0.000
Superior Temporal Gyrus	39	R	48	-55	25	0.000
- small volume-corrected peak						0.000*
Inferior Parietal Lobule	40	R	69	-31	28	0.000
- small volume-corrected peak						0.000*
Postcentral Gyrus	2	R	45	-25	31	0.000
Middle Temporal Gyrus	39	L	-42	-61	25	0.000
Inferior Parietal Lobule	40	L	-57	-28	25	0.000
Inferior Parietal Lobule	40	L	-48	-34	28	0.000
Precuneus	19	R	33	-79	34	0.001
- small volume-corrected peak						0.044*
Precuneus	7	L	-21	-79	49	0.002

C

Congenitally Blind Participants (N = 1) (WB)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
vOICe Noise Pause Detection [Post – Pre]						
No Activation						
vOICe Distract Counting [Post – Pre]						
Insula	13	L	-39	-4	1	0.001
Lentiform Nucleus		L	-21	2	4	0.002
Lentiform Nucleus		L	-30	-1	7	0.003
Posterior Cingulate	30	R	3	-49	19	0.001
Culmen		R	15	-40	-8	0.001
Lingual Gyrus	19	R	15	-49	-2	0.001
- small volume-corrected peak						0.024*
Superior Frontal Gyrus	6	R	12	23	64	0.001
Superior Frontal Gyrus	6	R	9	11	64	0.004
Clastrum		L	-33	-22	4	0.001
Middle Frontal Gyrus	11	L	-36	35	-11	0.003
Culmen		L	-9	-31	-14	0.003
Parahippocampal Gyrus		L	-33	-4	-20	0.005
Star Trek Sound Pause Detection [Post – Pre]						
Subcallosal Gyrus	34	R	12	2	-11	0.001
Subthalamic Nucleus,		R	12	-13	-5	0.002
Midbrain						
Beach Sound Pause Detection [Post – Pre]						
No Activation						

D

Congenitally Blind Participants (N = 1) (SB)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
vOICe Noise Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	R	48	-31	25	0.000
<i>- small volume-corrected peak</i>						0.000*
Supramarginal Gyrus	40	L	-42	-37	31	0.000
Cuneus	18	R	3	-97	22	0.000
<i>- small volume-corrected peak</i>						0.000*
Middle Frontal Gyrus	8	L	-42	23	43	0.000
Middle Frontal Gyrus	8	L	-30	35	46	0.000
Superior Frontal Gyrus	6	R	18	29	55	0.000
Superior Parietal Lobule	7	L	-33	-52	64	0.000
Postcentral Gyrus	7	R	9	-49	64	0.006
Cingulate Gyrus	32	L	-21	2	34	0.000
vOICe Distract Counting [Post – Pre]						
Cingulate Gyrus	24	L	0	11	31	0.000
Middle Frontal Gyrus	9	L	-27	20	34	0.001
Insula	13	R	51	-19	22	0.003
Star Trek Sound Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	L	-42	-34	28	0.000
Insula	13	L	-51	-19	25	0.000
Inferior Parietal Lobule	40	L	-48	-46	40	0.000
Inferior Parietal Lobule	40	R	51	-25	25	0.001
Precuneus	31	L	-15	-43	34	0.001
Beach Sound Pause Detection [Post – Pre]						
Superior Frontal Gyrus	6	L	-18	11	67	0.000
Superior Parietal Lobule	7	L	-27	-58	67	0.000

Congenitally Blind Participants ($N = 1$) (SB) Continued						
Region	BA	Side	x	y	z	p_{uncorr}
Insula	13	R	48	-25	22	0.001
Postcentral Gyrus	3	L	-45	-19	64	0.001
Inferior Parietal Lobule	40	L	-51	-49	43	0.002
Inferior Parietal Lobule	40	L	-39	-34	31	0.002
Cingulate Gyrus	24	R	3	-10	34	0.003

Table 4.4. fMRI data: Post – pre training blind and severe low-vision participants. Select imaging results for severe low-vision participant ($N = 1$), Panel A, a late blind participant ($N = 1$), Panel B, and two congenitally blind participants ($N = 2$), Panel C and D, when comparing post-vOICE-training scan and the pre-vOICE-training scan. For the late blind participant, only the top 15 clusters of activation are presented in Panel B; a full list is in Appendix D, Table B. All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

Sighted Participant Covariate Analyses of Imaging Data

Following the final fMRI imaging session, participants filled out a post-experiment questionnaire to identify the role of visualization, the completeness of the distraction in the distraction task, and other relevant parameters. This was meant to see whether the subjective post-hoc report of visualization and/or attentiveness was correlated with fMRI activation in the visual cortex. The survey is presented in full in Appendix C. The sighted participant results ($N = 10$) from four of the questions from this questionnaire are presented in Figure 4.14. Figure 4.14A shows that the number of sighted participants that recognized the vOICE noise as the vOICE device dramatically increased from the first fMRI session (pre-training) to the second session (post-training). Figure 4.14B shows that most participants in the pre-training and post-training fMRI scans were not distracted from the task of counting by the playing of the vOICE sound in the background. The final two plots, Figure 4.14C and 4.14D, indicate that the number of people imagining numbers in the counting task or visual scenes in the familiar sound pause task did not dramatically change between the pre- and post-training sessions.

The data presented in Figure 4.14 (experiment questionnaire) and Figure 4.08 (localization accuracy with training) can be used for fMRI covariate analyses. A covariate analysis (details in methods) determines whether a neural activation correlates across participants with a behavioral metric or the subjective post-hoc reports. The behavioral metrics and subjective post-hoc reports used here will include the visual imaginings during the distraction task to determine whether any of the visual activation in the distraction post-pre analysis is due to visualization, and the performance metrics at localization (slope and intercept) to determine whether any vOICE learning correlates

with the visual activation in the distraction task. Both of these covariates are designed to narrow down the possible origins of the visual activation in the distraction task, ideally showing that visualization did not play a role, and the vOICE learning did. With this correlative tie between the visual activation in the distraction task and the vOICE learning, it can more positively be stated that crossmodal plasticity with vOICE is automatic (*i.e.*, can occur without attention).

The results for the covariate analysis on the distraction vOICE results are presented in Table 4.5. The first covariate tested was to disprove the visualization hypothesis, that the activation in distraction task was due to visual imaginings of shapes or numbers. The covariate for this analysis was generated from the data in Figure 4.14C by making imagined numbers response = 1, not imagining numbers response = 0, and summing across the pre- and post-training fMRI sessions (*i.e.*, the max number was 2 if participant imagined in both sessions, and the minimum was 0 if the participant imagined in no sessions). The analysis indicates that no neural activation correlated with the imagining number behavioral metric. In addition, the number of participants that imagined numbers decreased from the pre-training to the post-training sessions (Figure 4.14C). Therefore, combining the covariate analysis and the decreasing number of participants with visualization from pre to post, it is unlikely that the visual activation in the distraction counting task was due to visualization.

The visualization covariate result is the most important and valuable to this study. This null result for visualization in the vOICE distraction counting task (Table 4.5) indicates that the visual activation in the distraction task is not likely due to visualization of numbers and shapes. Therefore, the visual activation is likely from crossmodal

plasticity engendered automatically from vOICE training. In addition, the visualization covariate result strengthens the case for automatic processing with vOICE in early visual regions of cortex (*i.e.*, BA 18 and 19).

The two other covariates performed on the distraction vOICE [Post – Pre] results used the vOICE training performance at the localization task. It is useful to note that localization performance is not a broad metric of vOICE interpretation ability, nor is the distraction task in fMRI using localization. Therefore, any neural activations that correlate with the vOICE localization covariates may be interesting, but the lack of a correlation between localization performance and visual activation from a vOICE counting distraction task would not diminish the vOICE fMRI results. It is particularly important to be aware of this for the vOICE device, because participants perform at different levels for different vOICE tasks. For example, a participant that is excellent at vOICE localization may be poor at recognition with vOICE. Therefore, for a vOICE performance metric to be a valuable covariate, it should be as close as possible to the vOICE task in the fMRI scanner. Since vOICE localization and the vOICE counting distraction task are not that behaviorally similar, it diminishes the value of this vOICE covariate fMRI result. This qualification to covariate correlation analyses will be revisited in more detail later.

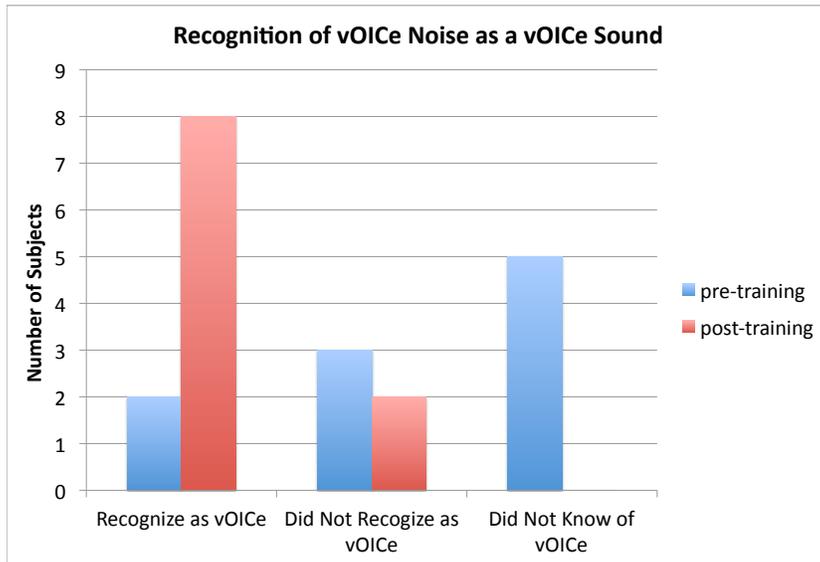
The first vOICE covariate is the performance improvement (slope) for the localization inaccuracy vs. training time plot (Figure 4.08 and 4.09). Since the y-axis is a measure of inaccuracy, a smaller slope (*i.e.*, more negative) means a better learning rate. Therefore, the slope of each participants data was multiplied by -1 to invert the data, making the larger slope values represent the best performing participants. This slope

covariate generated several neural regions that correlated with the slope (Table 4.5), including BA 6, 8, 24, and 32. These neural regions have been known to be involved with motor functions, auditory imagery, language, memory, executive functions and visuospatial attention. It is likely that these regions engaged in several of these functions during the counting distraction task. This frontal lobe region activation may correlate with participant improvement at the localization task, because participants that improved the most used a cognitive, top-down strategy in learning, therefore engaging pre-frontal regions more vigorously than the participants that did not improve as much. Since improvement (slope) and initial performance (intercept) are anticorrelated (Figure 4.10), it is also conceivable that the participants that did not improve as much were *better at the beginning of training*, and therefore engaged in a more automatic, perceptual strategy (with less frontal neural activity) based on crossmodal correspondences.

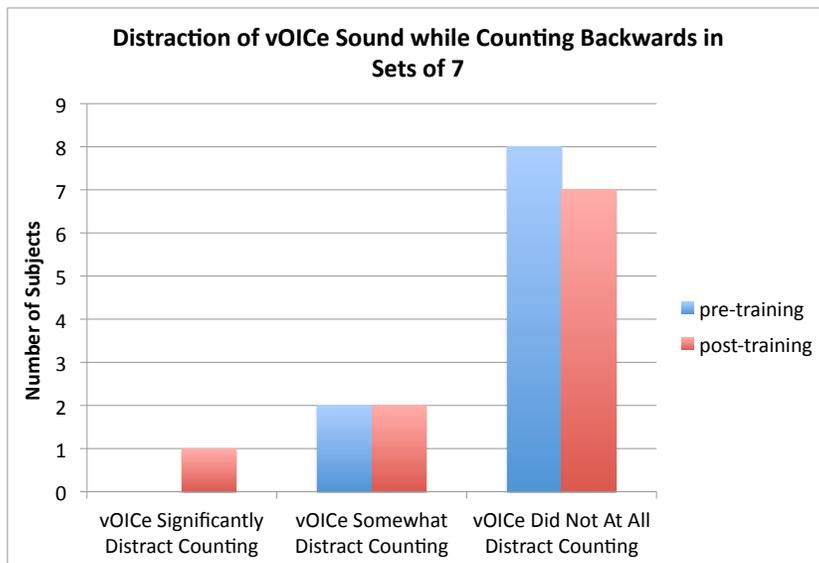
The second covariate based on localization vOICe data uses the beginning performance (intercept) of the localization performance *vs.* training time. The localization data is plotted as inaccuracy *vs.* training time (smaller values = more accurate localization). Therefore, to make the largest values the most accurate, all intercept values were multiplied by -1 (larger values = most accurate). No neural activation correlated with initial performance at vOICe for the [Post – Pre] distract counting task (Table 4.5). It is logical that this would be true, as the contrast compares the post-training scan with the pre-training scan, therefore identifying the changes due to training, whereas the covariate is for the initial performance and not training changes. While this covariate is less valuable to the vOICe distraction [Post – Pre] contrast, it will be more relevant when

used with later contrasts in Chapter 5 that compare tasks within one fMRI session (*i.e.* only before training, or only after training).

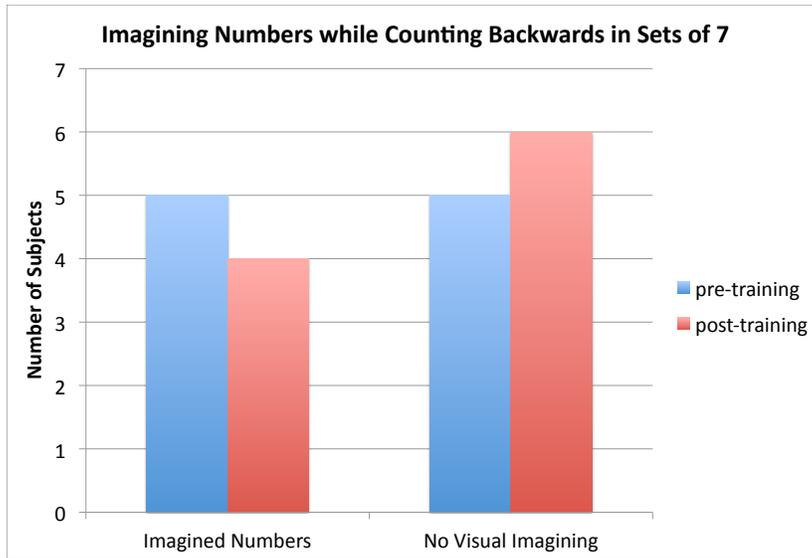
A



B



C



D

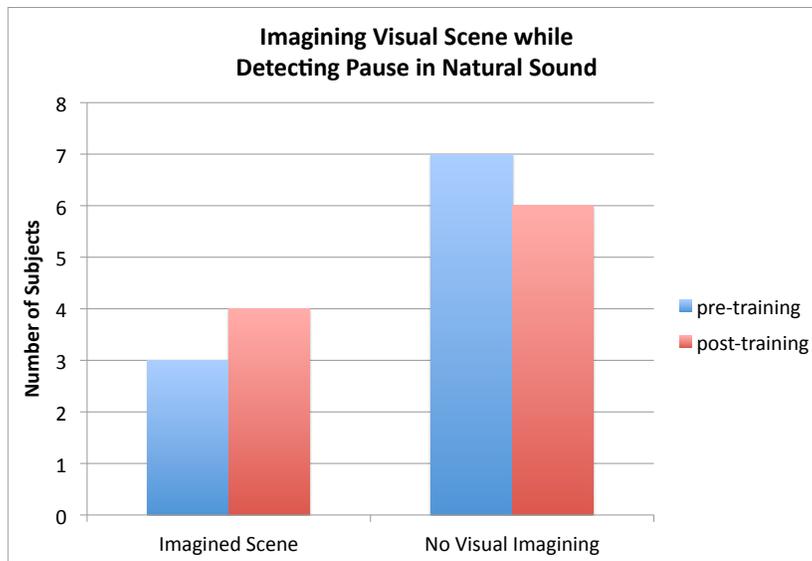


Figure 4.14. Post-experiment Questionnaire Results. Following the post-training fMRI scan, all participants filled out a questionnaire (Appendix C). This figure plots the responses to select questions in that questionnaire for the 10 sighted participants.

Sighted Participants ($N = 10$)						
Region	BA	Side	x	y	z	p_{uncorr}
Distract Counting [Post – Pre] Visualization Covariate						
No Activation						
Distract Counting [Post – Pre] Localization Slope Covariate						
Sub-Gyral	6	L	-15	-4	52	0.001
Cingulate Gyrus	24	L	-15	-4	43	0.001
Middle Frontal Gyrus	6	L	-30	-10	43	0.001
Superior Frontal Gyrus	8	R	21	26	46	0.001
Medial Frontal Gyrus	6	L	-9	-25	67	0.001
Medial Frontal Gyrus	6	R	6	-25	67	0.003
Medial Frontal Gyrus	32	R	6	8	46	0.003
Medial Frontal Gyrus	6	L	0	-4	49	0.003
Distract Counting [Post – Pre] Localization Intercept Covariate						
No Activation						

Table 4.5. fMRI covariate data: Post – pre training vOICE noise distract sighted participants. Three covariates for vOICE Noise Distract task are displayed in this table: one for visualization, and two based on vOICE training performance. Details on the processing of covariates is in the methods section and the results section of Chapter 4. The neural activation shown for vOICE Noise Distract [Post – Pre] correlates with the performance of the covariate listed, indicating that the covariate may have played a role in generating the neural activation listed. All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold. Brodmann Area localization was performed

on the talarach client for nearest grey matter. Any clusters without nearest grey matter within ± 5 mm are not included.

Sighted Participant Visual Suppression Following vOICe Training

All previous Chapter 4 fMRI contrasts have investigated the impact of vOICe training on crossmodal interaction and plasticity. The results of these contrasts indicated that in the sighted participants ($N = 10$), the severely impaired ($N = 1$), and most blind participants ($N = 2$) there was automatic activation of multimodal or visual regions to vOICe stimuli. It is also interesting to investigate whether this crossmodal plasticity had an impact on traditional visual perception in the sighted participants. It is possible that the vOICe-based new crossmodal connectivity enhances the effectiveness of processing in the visual cortex. In particular, the blindfolding of sighted individuals has been known to increase visual region excitability (Boroojerdi et al., 2000). As the sighted participants were blindfolded for 5 hours during our vOICe training, this is a possible outcome. However, it may also be possible, as an alternative, that auditory (or crossmodal) connections to visual cortex are in competition with visual connections to visual cortex. Therefore, when the crossmodal influence on visual cortex is increased by the vOICe training, the visual dominance of visual cortex may be slightly weakened. In support of this hypothesis, a study by Rauschecker and Korte used visual deprivation on cats to induce neuron sensitivity to auditory stimuli and to decrease sensitivity to visual stimuli in the anterior ectosylvian (AE) cortex known for visual processing alone (Rauschecker & Korte, 1993). They conjectured that because the visual response was reduced when the crossmodal activations of AE were increased, that the two types of input were in competition for dominance.

We tested for this suppression or strengthening of visual activation of visual cortex with a simple visual task performed before and after vOICe training in sighted

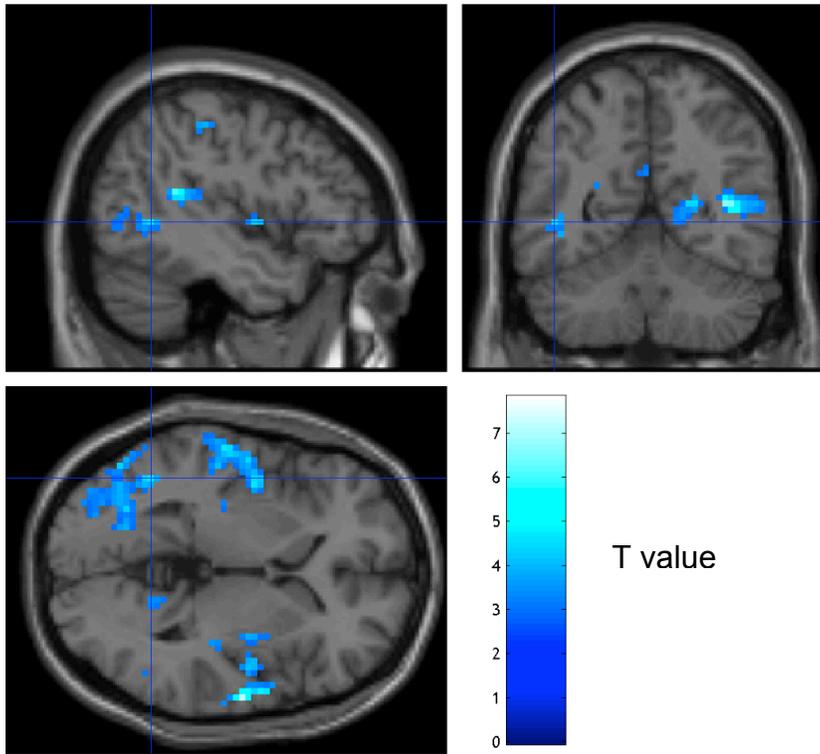
participants. Participants were asked to view a white noise visual image and detect whether it paused during its presentation, and then respond with a button press. The imaging contrast of this task (vision noise pause detection) was activation following vOICe training subtracted from activation before vOICe training to determine whether any visual regions were less active after the vOICe training compared to before (Table 4.6 and Figure 4.15). The results indicate that several visual and multisensory regions' (such as BA 19, and 40) activity were suppressed significantly following the vOICe training. These regions of visual suppression are similar to the regions crossmodally activated by vOICe stimuli in the same sighted participants (Table 4.1 and 4.2) and in the blind participants (Table 4.4). In other words, two processes (visual and crossmodal) are likely competing for the activation of same visual and multimodal regions; as crossmodal plasticity is strengthened (as seen in the vOICe tasks), traditional vision is weakened (as seen in the visual noise task).

It is important to control for the possibility that not visual suppression but rather fatigue or inattention is causing a reduction in visual activation following vOICe training. Certainly neural fatigue or diminished interest could also cause a decrease in visual activation in the second session (following vOICe training) in comparison to the first session (before vOICe training). To control for these possibilities, we used a covariate of performance of participants during vOICe training to determine whether any regions of suppression correlated with the vOICe performance. If suppression in the regions of interest correlate with vOICe performance, then that would indicate that the suppression is likely tied to vOICe training rather than other spurious factors.

The first vOICe performance covariate tested is the amount of improvement at the localization task (Figure 4.08) or the negative of the slope in the localization inaccuracy curve (covariate detailed in methods p. 144, and in previous section p. 178-186). No regions of neural activation for vision noise pause detection [Pre – Post] correlated with the vOICe improvement covariate.

The second covariate is the initial performance at the vOICe localization task (Figure 4.08), or the negative of the y-intercept in the localization inaccuracy curve (covariate detailed in methods p. 144, and in previous section p. 178-186). Many visual and multimodal regions that were similar to the regions of visual suppression in the vision noise contrast, did indeed correlate with this second initial vOICe performance covariate (BA 40, 17, and 18; BA 17 is significant with small volume correction; Table 4.7). Therefore, this result means that the visual and multimodal suppression is likely tied to the vOICe training between the sessions rather than changes in arousal or neural fatigue. The initial performance correlation with visual suppression may demonstrate that the initial individual crossmodal strength plays an important role in shaping the neural dynamics of crossmodal plasticity, and therefore visual suppression. Improvement at the vOICe task may be less useful than the initial performance as a covariate with visual suppression, because improvement at localization with vOICe is limited by the performance ceiling, whereas initial performance is not.

A



B

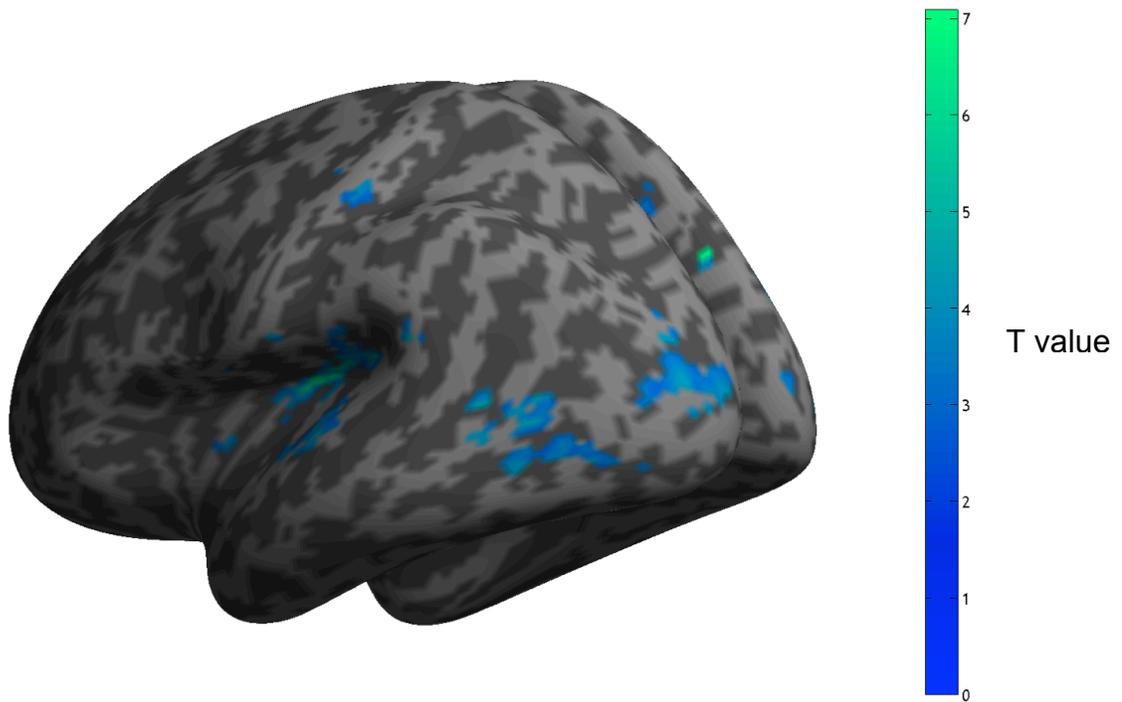


Figure 4.15. fMRI data: Pre – post training vision noise pause detection task in sighted participants. The neural imaging result is displayed for pre-vOICe-training in contrast to post-vOICe-training for the vision-noise pause detection task in sighted participants ($N = 10$). Imaging data presented is $p < 0.009$ uncorrected and clusters of 10 voxels or more; further correction for multiple comparisons is shown in Table 4.6. Methods for fMRI data display are in the Chapter 4 methods. Activation is shown in blue to indicate that activation with a [Pre – Post] contrast is in fact representing a suppression of activation if the typical [Post – Pre] contrast was used (as was used in all other contrasts in Chapter 4). Therefore, relative to all other Chapter 4 figures, which present visual activation, Figure 4.15 represents visual suppression. The detailed description of methods for fMRI data display are in the Chapter 4 methods.

Sighted Participants (N = 10)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
Vision Noise Pause Detection [Pre – Post]						
Superior Temporal Gyrus	22	R	66	-10	4	0.000
Insula	13	R	36	-19	19	0.000
Precentral Gyrus	6	R	54	-1	13	0.000
Superior Temporal Gyrus	13	L	-42	-43	16	0.000
Insula	13	L	-30	-28	16	0.000
Superior Temporal Gyrus		L	-60	-19	1	0.001
Precuneus	7	R	21	-70	31	0.000
Superior Temporal Gyrus	22	R	39	-55	10	0.000
Posterior Cingulate	30	R	27	-70	13	0.000
Superior Temporal Gyrus	22	L	-63	-46	19	0.000
Inferior Parietal Lobule	40	R	57	-31	46	0.000
- small volume-corrected peak						0.020*
Precentral Gyrus	4	R	51	-7	46	0.001
Postcentral Gyrus	2	R	54	-25	55	0.001
Paracentral Lobule	31	R	9	-10	43	0.000
Paracentral Lobule	31	L	0	-10	46	0.001
Cingulate Gyrus	24	R	9	-1	46	0.002
Precuneus	31	L	-3	-61	28	0.001
Posterior Cingulate	23	L	0	-46	25	0.006
Inferior Temporal Gyrus	19	L	-45	-55	1	0.001
- small volume-corrected peak						0.035*
Middle Occipital Gyrus	19	L	-30	-88	10	0.001
Lingual Gyrus		L	-30	-73	4	0.001
Inferior Parietal Lobule	40	L	-48	-28	49	0.001

Sighted Participants ($N = 10$) Continued						
Region	BA	Side	x	y	z	p_{uncorr}
Superior Temporal Gyrus	22	R	48	-7	1	0.002
Precuneus	31	L	-18	-73	28	0.002
Cuneus	18	L	-12	-76	22	0.002
- small volume-corrected peak						0.076*
Precuneus	7	R	9	-67	43	0.003

Table 4.6. fMRI data: sighted participants' vision noise pause detection [Pre – Post]. Imaging results for sighted participants ($N = 10$) when comparing pre-vOICE-training scan and the post-vOICE-training scan for visual noise pause detection task. All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

Sighted Participants (N = 10)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
Vision Noise Pause Detection [Pre – Post] Localization Slope Covariate						
No Activation						
Vision Noise Pause Detection [Pre – Post] Localization Intercept Covariate						
Insula	13	R	39	-13	13	0.000
Superior Temporal Gyrus	22	R	51	-10	1	0.000
Superior Temporal Gyrus	22	R	57	-1	1	0.001
Superior Temporal Gyrus	22	R	60	-34	13	0.000
Superior Temporal Gyrus	41	R	48	-31	13	0.000
Inferior Parietal Lobule	40	R	69	-37	28	0.001
- small volume-corrected peak						0.064*
Superior Temporal Gyrus	41	L	-48	-34	10	0.000
Superior Temporal Gyrus	22	L	-45	-28	1	0.001
Superior Frontal Gyrus	6	L	0	11	67	0.000
Superior Frontal Gyrus	6	R	9	5	67	0.003
Superior Temporal Gyrus	41	L	-57	-19	4	0.000
Transverse Temporal Gyrus	42	L	-63	-16	10	0.001
Cuneus	17	L	-12	-94	4	0.000
- small volume-corrected peak						0.029*
Thalamus, Pulvinar		R	15	-28	10	0.003
Superior Frontal Gyrus	6	L	-21	-1	67	0.001
Cuneus	17	R	18	-82	7	0.002
Middle Occipital Gyrus	18	R	30	-79	4	0.008
- small volume-corrected peak						0.167*

Table 4.7. fMRI covariate data: Pre – post training vision noise in sighted participants.

Two covariates for the vision noise pause task are displayed in this table; both are based

on vOICe training performance. Details on the processing of covariates is in the methods section and the results section of Chapter 4. The neural activation shown for vision noise pause detection [Pre – Post] correlates with the performance of the covariate listed, indicating that the covariate may have played a role in generating the neural activation listed. All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold. The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included

Discussion

Sensory substitution interpretation and neural processing has been presumed to be serial, cognitive, and not automatic (detailed in Chapter 1). This fMRI study of the automaticity of sensory substitution neural activation has dramatically altered this top-down theory of sensory substitution processing. Our results indicate that sensory substitution can be processed in visual cortical regions without attention or image structure (*i.e.*, white noise image used) in sighted and blind vOICe users. Imaging correlations with participant post-hoc reports show that automatic visual activation from vOICe (*i.e.* during a distraction task) is not likely due to visualization. Further, an interesting result of visual suppression during a visual pause detection task when comparing post-vOICe-training to pre-vOICe-training indicates that crossmodal and natural visual processing may be in competition for dominance in visual cortical regions.

vOICe in Comparison to Vision

The passive processing of sensory substitution in visual regions is similar to visual processing, which can occur at a diminished intensity for unattended objects (for details, see the beginning of Chapter 4). The identification of similarities between sensory substitution (SS) and vision has grown in recent years with the renewed interest in SS's potential. The similarities now include: Functional activation of visual regions (*i.e.*, activation of the FFA while recognizing faces), depth perception, recognition, constancies (detailed in Chapter 2), and causal activation of visual regions by SS (detailed in Chapter 1). The automatic activation of visual cortex can now be added to that list of similarities between SS and vision. Further, the identification of automatic visual processing of SS in visual neural regions indicates that there is automatic

perceptual processing of SS that may generate faster training and easier perceptual use. Perhaps the crossmodal mappings discussed in Chapter 3 are one way to tap into this automatic perceptual neural processing of SS.

Prism Adaptation and vOICe Learning

Studies focused on learning new hand-eye relationships via prisms indicate a similar learning pattern to the vOICe localization learning shown in this Chapter. When participants begin using the prism glasses with a shifted or rotated visual transformation, their performance deteriorates due to the inaccuracy of their existing perceptual processing in relation to rotated or shifted vision (Harris, 1965; Shimojo & Nakajima, 1981; von Helmholtz, 1925). However, as the participants use the glasses, their localization and reach and grasp performance gradually improves. Occasionally, the visual perception with prism glasses alters neural processing such that participants no longer visually perceive the shift or rotation from the glasses, indicating adaptive perceptual changes following sensory-motor adaptation (although not all experiments report this perceptual change) (Linden, Kallenbach, Heinecke, Singer, & Goebel, 1999). Both of these patterns occur with vOICe perception; when participants start using the new auditory-visual encoding and camera, their performance is not near the optimal localization performance. Yet, as shown in Figure 4.08 and Figure 4.09, the performance improves to a ceiling based on the systems resolution and the hand-camera coordination. Although not shown in this study, other sensory substitution studies indicate that some blind users have a similar alteration in visual perception to the prism users. In particular, as detailed in Chapter 1, Ortiz *et al.* found that a fraction of blind participants trained on a sensory substitution device also had visual experiences of stimuli perceived with the

device (Ortiz, et al., 2011). Further, other expert blind participants have claimed to perceive extensive visual experiences with sensory substitution (Ward & Meijer, 2010) (for details, see Chapter 1). It makes sense that prism learning and SS learning would have these commonalities, as both are the learning of a new hand-eye (or camera) coordination as well as a new transformation algorithm (prism = shift or rotation in vision, sensory substitution = audition to vision). In general, both learning patterns are due to plasticity that adapts to the new unexpected changes in perception, enabling functional learning and rehabilitation.

Automatic vs. Cognitively Demanding Crossmodal Plasticity

It is also interesting to discuss whether the neural activation from perceptual processing of sensory substitution after training uses the same crossmodal interactions as the plasticity evident during demanding cognitive tasks with sensory substitution. As discussed in Chapter 1 and Chapter 3, several crossmodal connections exist before SS training, such as crossmodal correspondences or the connections generating the double flash illusion. These crossmodal interactions can generate visual activation in response to auditory stimulation (detailed in Chapter 1), and may be occurring via direct connections between visual and auditory regions or through indirect feedback connections. Crossmodal plasticity generated by sensory substitution can use these existing connections and modulate their strengths or create new connections. Alternatively, the SS's crossmodal interaction could be occurring in a multimodal region such as the superior temporal sulcus or the angular gyrus (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Beauchamp, Lee, Argall, & Martin, 2004; Ramachandran & Hubbard, 2003; Spence, 2011). Current cognitively-demanding sensory substitution crossmodal

plasticity has been postulated to occur through a top-down (feedback) neural network that includes primary sensory regions as well as multisensory cortical areas (Chapter 1, Figure 1.8). Which feedback neural network or feedforward connections are used is an open question for both the automatic and cognitively-demanding SS plasticity. For detailed comparisons of neural network architectures, DCM modeling of sensory substitution fMRI data is required (DCM details on p. 45). In the meantime, the close correspondence of the neural imaging results (early visual activation) in the automatic task (*i.e.* vOICE distract counting task) and other cognitively demanding tasks in the literature indicate that their neural networks likely have some similarities.

TOPOGRAPHIC MAPPING OF VOICE CROSSMODAL PLASTICITY

Introduction

Sensory substitution neural imaging studies have focused on the presence of crossmodal plasticity, and the functional association of the task and neural region activated. No sensory substitution study has investigated the mapping of visual space via sensory substitution to a spatiotopic or “retinotopic” map. Retinotopic mapping processes adjacent regions in visual space by neighboring regions of cortex. Vision has a retinotopic map (detailed below) that is based on the 2D spatial luminance-detection of the two retinas. The basic principle of the visual retinotopic mapping is the representation of the contralateral visual field in each hemisphere (*i.e.*, left visual field is processed by the right primary visual cortex, and vice versa). This chapter will focus on an fMRI experiment determining whether the same contralateral mapping of visual space occurs with vOICe spatial perception. This is an intriguing question, because in the A-V type of sensory-substitution device (such as the vOICe), the auditory inputs after appropriate training may systematically induce early visual cortical activation (as reviewed in Chapter 1 and 4). Yet, which neural pathway and/or multisensory plasticity enable it, is not very well understood. The empirical data answering the spatial mapping of this crossmodal plasticity will shed light on the underlying neural pathways responsible for multisensory plasticity with the SS device/training.

Investigation of the spatial mapping to neural activation of SS may show that like vision, it is retinotopically mapped. Visual mapping of space begins at the retinal level

(Figure 5.1). Each retina detects an inverted image of both the left and right visual field, which is then separated into left and right visual fields at the optic chiasm. The left visual field fibers merge from both eyes and exit right of the optic chiasm, and vice versa. The fibers then continue to the thalamus' lateral geniculate nucleus (LGN) and onto primary visual cortex in the occipital lobe. Primary visual cortex or V1 spatially maps the hemifield on the cortical surface, as is elegantly shown in Tootell *et al.*'s mapping of the macaque monkey (Figure 5.2) (Tootell, Silverman, Switkes, & De Valois, 1982). Tootell and colleagues used a c-labeled deoxy-d-glucose prior to exposure of the animal to the visual target pattern (Figure 5.2 A). The chemical label could then be used to stain recently activated neural cells and therefore show the pattern of activation on the cortex itself. Their results (Figure 5.2 B) show the incredible fidelity of the retinotopic map to the original image with the following modification – that is, the map is logarithmically magnified at the foveal region relative to the periphery. More recently, fMRI imaging has been used to generate detailed retinotopic maps in humans for not only V1 but also V2, and V3 among others (Kolster, Peeters, & Orban, 2010; M. Sereno *et al.*, 1995; M. I. Sereno, Pitzalis, & Martinez, 2001). It is useful to note that the sharpness of the contralateral mapping (*i.e.*, left hemisphere to right visual field) decreases as information progresses from V1 to higher visual cortices. Consequently, extrastriate regions represent an increasing amount of the ipsilateral visual field. For example, Tootell *et al.* determined in 1995 that MT neurons responded to visual stimuli up to 20 degrees into the ipsilateral receptive field (Tootell *et al.*, 1995).

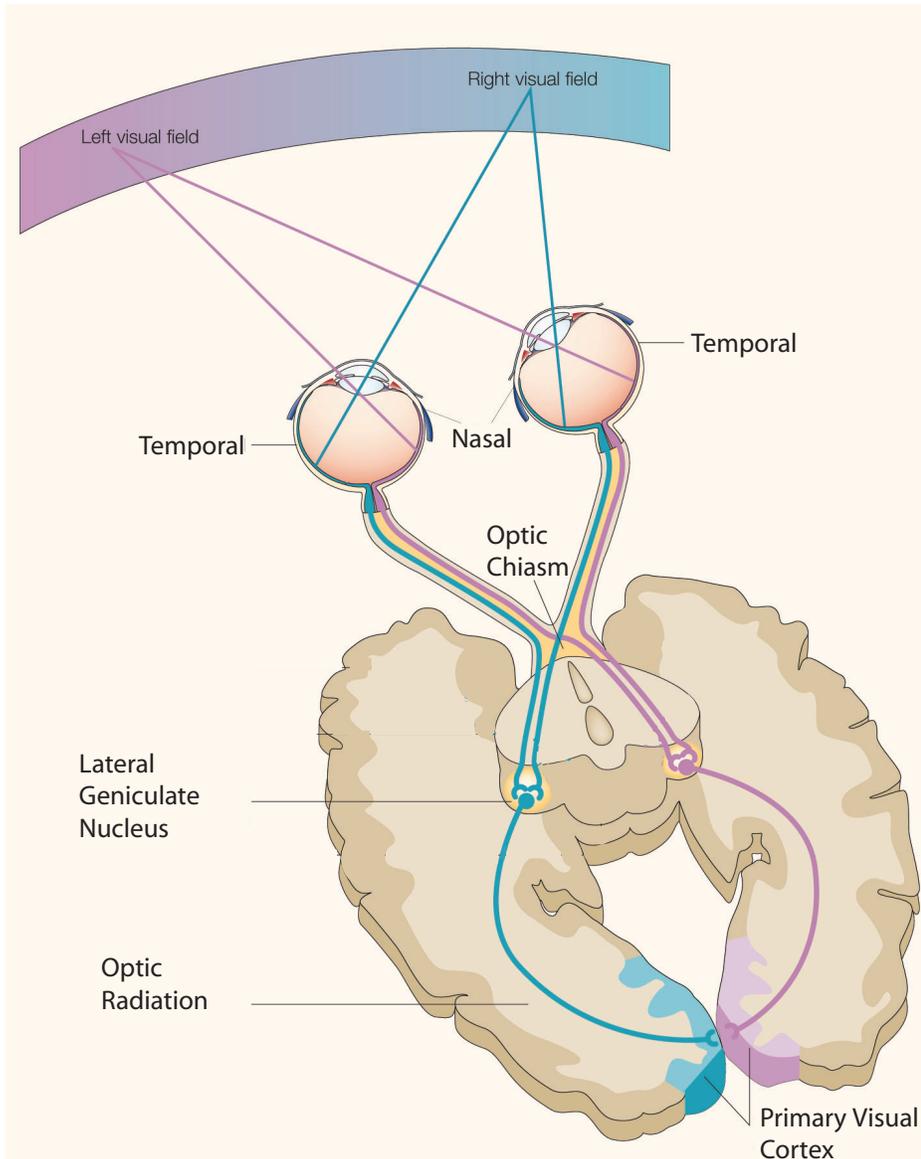


Figure 5.1. Diagram of the mapping of visual space to cortical activation. This diagram indicates the mapping of the left and right visual fields from the retina, through the optic chiasm and lateral geniculate nucleus to the primary visual cortex (V1). It is accomplished by the nasal part of each of the retinae projects to the contralateral, whereas the temporal part of it projects to the ipsilateral visual cortices (Hannula, Simons, & Cohen, 2005).

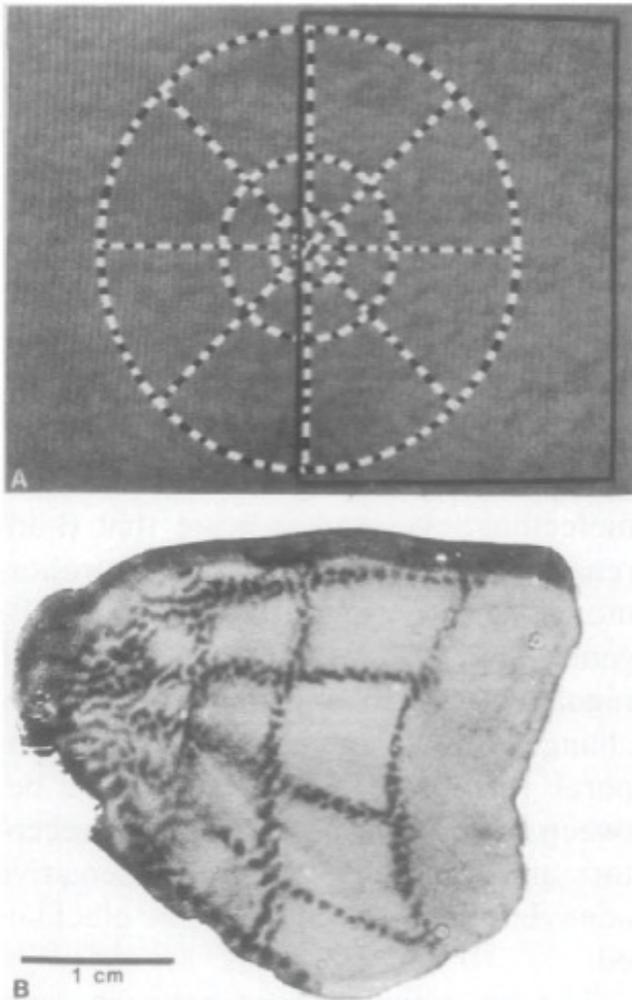


Figure 5.2. Retinotopic mapping of macaque striate cortex. Tootell and colleagues demonstrate the retinotopic mapping of visual space onto the primary visual cortex with a deoxyglucose analysis method (Tootell, et al., 1982). The image of visual cortex (image B) shows half of the pattern visually presented to the monkey (image A), indicating the mapping of half visual space to neural activation (indicated by dark patches of cortex) in the contralateral hemisphere and the mapping of neighboring spatial regions to adjacent regions of visual cortex. (Tootell, et al., 1982)

Visual retinotopic maps are plastic, and can be modified by visual deprivation. Several studies have investigated the implications to retinal mapping when a cortical lesion occurs, whether via stroke or surgical intervention. A patient with the loss of one visual field due to stroke in the left hemisphere of the occipital cortex was able to remap both visual fields onto the visual regions of the intact occipital lobe, including V1 (Henriksson, Raninen, Nasanen, Hyvarinen, & Vanni, 2007). Further, a participant with a right hemispherectomy (removal of the right hemisphere) due to epilepsy had visual activation in the left V3 and the left V5 in response to stimuli in the blind visual field (left side); thus the projection was plastically re-organized ipsilaterally (Bittar, Ptito, Faubert, Dumoulin, & Ptito, 1999). Interestingly, this right hemispherectomy participant also experienced blind sight (unconscious visual perception), whereas the two other hemispherectomy participants in the study did not have visual activation from their blind hemifield or blind sight.

Visual retinotopic maps can also be modified by altered visual perception via prism glasses (detailed prism discussion in Chapter 4 discussion). Sugita found that the primate visual cortex became sensitive to the ipsilateral visual field after wearing left-right reversing prisms for one and a half months (Sugita, 1996). Further, Miyauchi and colleagues used fMRI imaging in humans wearing left-right reversing glasses to show that V1 and extrastriate visual regions became sensitive to ipsilateral visual stimuli (Miyauchi et al., 2004). These remapping results are valuable indicators of the plasticity of spatial maps in early visual regions.

Other investigations have studied the impact of retinal diseases on cortical maps, in particular investigating whether remapping occurs in the de-afferented cortex. Baseler

et al. studied the responsiveness of the cortical region that represents the all-cone foveola in congenital rod monochromats (colorblind people with nearly no cone receptor function) (Baseler et al., 2002). Baseler and colleagues determined that in rod-monochromats, remapping occurred in the foveola cortical region. The foveola now responded to rod-dominated retinal regions. The reorganization of de-afferented cortex of late-onset retinal diseases is less clear. In particular, Baseler and colleagues argued in 2011 that remapping does not occur in humans with bilateral central vision damage from Age-Related Macular Degeneration (AMD) (Baseler et al., 2011).

Even blind individuals have been shown to have spatiotopic maps of perception in visual regions. It was reported that a visually-impaired participant was able to activate normally foveal visual areas by Braille reading, while normally peripheral visual areas were activated by his remaining low vision (Cheung, Fang, He, & Legge, 2009). Further, Milne *et al.* were able to map azimuth of echo-locations on the visual cortex in an early blind echolocation expert in a way that is similar to the visual spatiotopic map (Milne, Goodale, & Thaler, 2013). Therefore, it is plausible that sensory substitution could generate a spatiotopic map in visual cortex.

This chapter will report experiments that investigate whether crossmodal plasticity with vOICE can be spatiotopically organized. The main fMRI task before and after training on vOICE will ask participants to localize a dot on the left or right, with the dots conveyed via vOICE sounds or via images. The mapping of visual space via vOICE to visual activation will then be determined by the comparison of the neural activation from the left dot and the right dot. Both sighted and blind individuals will participate in

the task, and thereby indicate whether the spatial mapping is different among these participant groups.

Methods

Participants

Participant information is detailed in Chapter 4 methods (p. 130). The same participants and scan sessions were used for Chapter 5 fMRI data collection as were used in Chapter 4. Chapter 5 analyzes the fMRI data results for two localization tasks not detailed in Chapter 4.

Experiment Design

The Chapter 4 methods describe the experimental design for both Chapter 4 and 5. Figure 4.01 also details the experimental layout for both Chapter 4 and 5.

vOICe Training Procedure

Participant training procedure on the vOICe device is explained in Chapter 4 methods, and in Appendix B part 1.

fMRI Tasks

Overview

Six separate tasks were performed in each fMRI scanning session. The four tasks relevant to the automaticity of vOICe processing are described in Chapter 4's methods. The remaining 2 tasks relevant to the mapping of vOICe from visual space to visual activation (*i.e.*, Chapter 5) are explained below.

vOICe Dot Localization

To perform the vOICe localization task, participants are asked to fixate on a cross in the center of the field of view, and listen to an image of white dot encoded into sound with vOICe play twice (Figure 5.3). The white dot can be located in the left visual field (*i.e.*, on the left side) or in the right visual field (*i.e.*, on the right side). Participants are asked to press 1 if the dot is located on the left, and 2 if the dot is located on the right. Participants after training are told that the sound is vOICe, and that this task is like the localization performed with vOICe during training. Before training, participants are typically told to press 1 if they hear a high-pitched sound on the left, and to press 2 if they hear it on the right. The vOICe sounds paired with the correct responses are also indicated before the participants start the task in both pre-training and post-training sessions. The participant's eye movements were recorded in both sessions to verify that participants move their gaze minimally, therefore not significantly modifying their spatial frame of reference. Participants performed 100 total trials of vOICe localization before and after training; 50 trials for the left-sided dot, and 50 trials for the right-sided dot, in randomized order.

Vision Dot Localization

To perform the visual localization task, participants were asked to fix their gaze on a central cross and locate a white dot presented on the left or right of the image center (Figure 5.4). Participants responded by pressing 1 if the white dot was on the left, and 2 if the white dot was on the right. The participant's eye movements were recorded in both sessions to verify that participants move their gaze minimally, therefore not significantly modifying their spatial frame of reference. Participants performed 100 total trials of

visual localization before and after training; 50 trials for the left-sided dot, and 50 trials for the right-sided dot, in randomized order.

Blind Participant Tasks

Blind participants performed the vOICE dot localization task, but not the vision dot localization task. Instructions for the vOICE dot localization task were read aloud by the Macintosh Computer Speech utility and recorded by QuickTime into an audio mov file. These mov files were converted into wav files, and loaded into MATLAB to be played at the beginning of the experiment. Eye movements were not recorded for blind participants. All other elements of the experimental design were the same for the blind participants, including the vOICE training.

fMRI Data Acquisition, Preprocessing, and Postprocessing (Statistical Analysis)

fMRI data collection parameters, and data pre- and post-processing details are in the Chapter 4 methods.

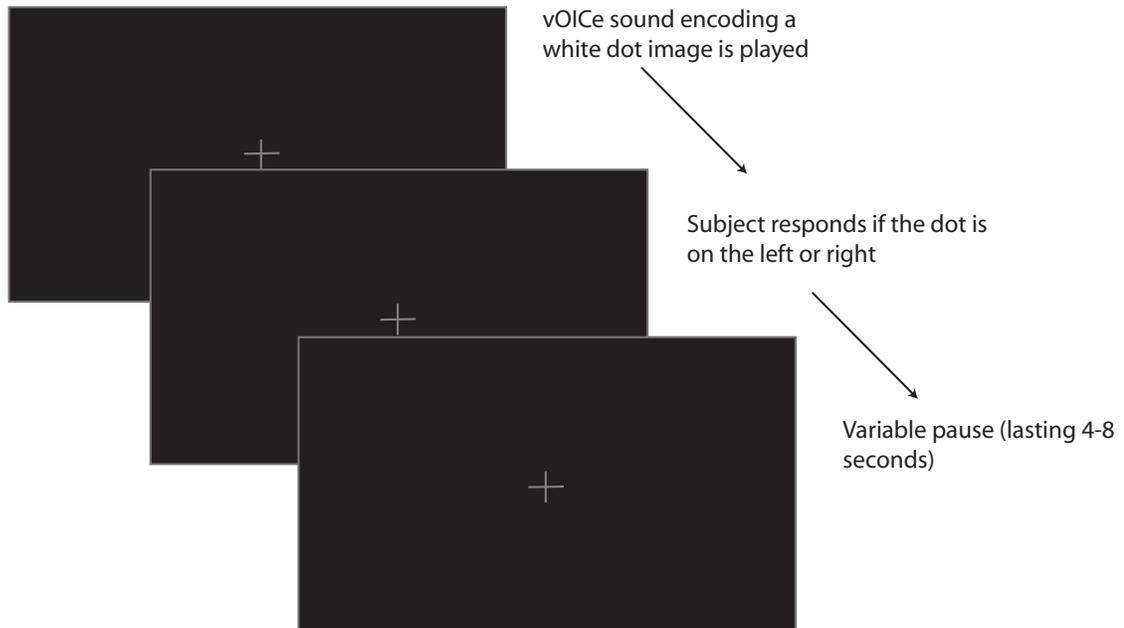


Figure 5.3. fMRI experiment diagram of the vOICE localization task. Participants localized a white dot on black background encoded into vOICE on the left or right. They responded after the sound finished if the dot was on the left by pressing 1, and the right by pressing 2. One hundred localization trials were performed, with 50 left dot trials and 50 right dot trials.

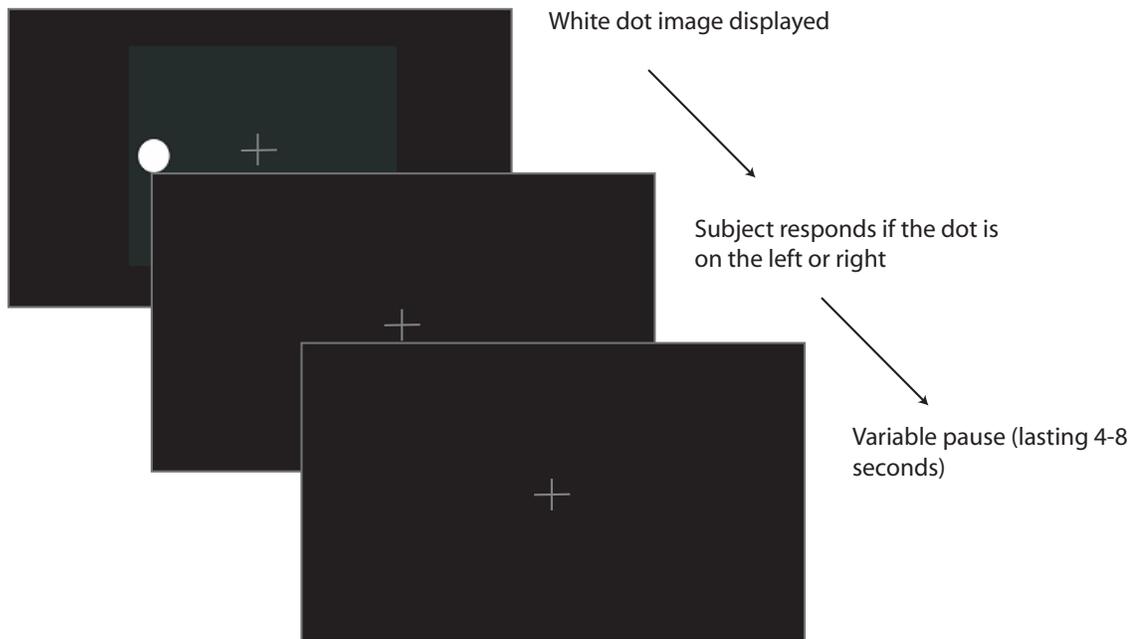


Figure 5.4. fMRI experiment diagram of the vision localization task. Participants localized a white dot on black background on the left or right with vision. They responded after the image disappeared if the dot was on the left by pressing 1, and the right by pressing 2. One hundred localization trials were performed, with 50 left dot trials and 50 right dot trials.

Results

Sighted Participant Neural Imaging Results

The contrast of vOICe localization [Right – Left] post-training and [Left – Right] post-training were used to investigate the spatial mapping of vOICe perception in 10 sighted participants (Note: [Right – Left location] and [Right – Left] will be used to indicate that the left dot location scans were subtracted from the right dot location scans and vice versa; the word(s) *post-training* after the brackets indicates that all the scans were all derived from the post-vOICe-training scan session). If a contralateral mapping exists, [Right – Left] will generate visual activation in the left hemisphere and [Left – Right] in the right hemisphere. Figure 5.6 and Table 5.1 show the results for this contrast. For the [Right – Left] vOICe contrast, significant activation was found in Brodmann Areas (BA) 19, 39, and 22 in the left hemisphere among other regions. A small volume correction for BA 19 yielded a pvalue less than 0.05, when a sphere of 10 mm radius was used (Table 5.1). For the contrast of [Left – Right] post training vOICe (Table 5.1), only the cingulate gyrus was activated. Therefore, the sighted participants had a contralateral mapping from a *right* dot in visual space to visual activation in *left* hemisphere of the brain. This contralateral mapping from space to visual neural activation via vOICe sound mimics the contralateral mapping in traditional visual perception.

In order to have a direct comparison of vOICe localization to vision localization, a vision control task was also performed by participants before and after training on the vOICe device. The vision contrast of [Right – Left] post-training and [Left – Right] post-training are shown in Figure 5.6 and Table 5.2. The [Right – Left] post-training in vision

had significant activation in BA 19 in the left hemisphere and BA 39 in the right hemisphere among other regions. A small volume correction of BA 19 and BA 39 yielded p-values below 0.05 for a sphere of 10 mm radius (Table 5.2). Nearly all of the neural activation in early visual regions (*i.e.* BA 17, 18, or 19) occurred in the left hemisphere, therefore indicating a largely contralateral mapping from visual space to neural activation (as expected). In a similar pattern to the vOICe localization results, the vision [Left – Right] post-training did not have any significant activation in visual or any other brain regions. This strange dominance of the right over the left visual field visual activation could have several possible causes. The repetition and simplicity of the task that could decrease visual activation (particularly because the scans studied above were in the second fMRI session) therefore make inter-hemisphere differences more apparent. It is also possible that vOICe training reduced the strength of visual activation via competition with the crossmodal visual activation as proposed in Chapter 4.

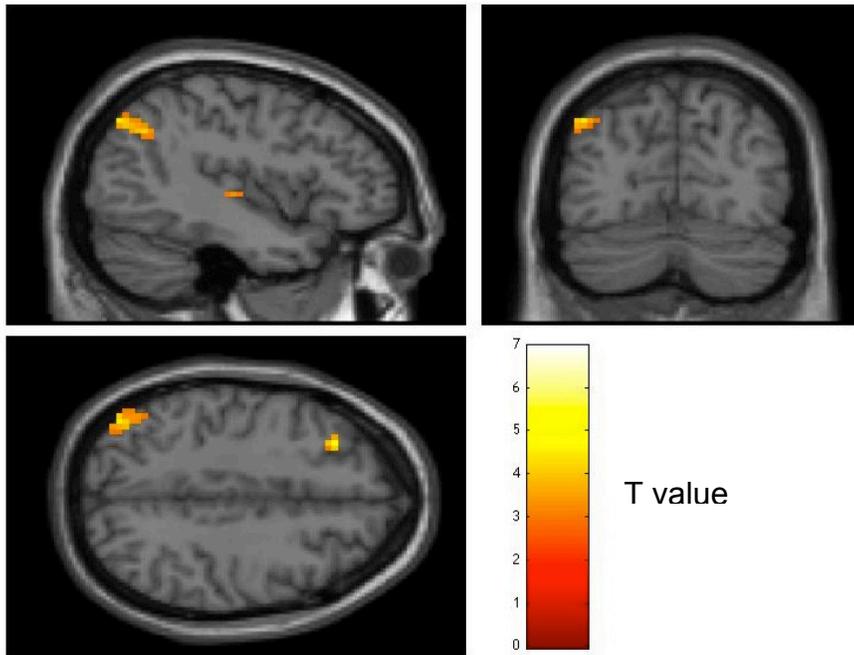
Sighted Participant Neural Imaging Correlation with Subjective Reports

All participants filled out a questionnaire following the experiment (full questionnaire in Appendix C). The results for a few questions relevant to the localization experiment are plotted in Figure 5.7. The first plot indicates that most participants did not visually imagine the dot while performing the localization task with the vOICe, and those that did have visual imagining only had them following the sound (Figure 5.7A). The scans used for the fMRI contrasts are only during the sound duration; therefore, any visual imaginings following the sound are not relevant for the fMRI analysis of the vOICe localization task. Therefore, this questionnaire data indicates that visualization did not play an important role in the generation of visual activation with vOICe localization.

The second plot presented (Figure 5.7B) shows that all sighted participants attempted to fixate their gaze during the vOICe localization task. This is particularly important because a wandering gaze is not only distracting, but may also alter the participant's visual frame of reference while performing the vOICe localization task, and therefore alter their spatial mapping of vOICe. This result critically shows that any visual wandering was minimized by active participant fixation during the localization task.

An fMRI covariate analysis was used to further tie the visual activation during the localization task with vOICe via vOICe performance during training. Two covariates were used for functional vOICe localization performance during training: The slope (improvement) and intercept (initial performance) of the vOICe localization inaccuracy *vs.* training time plot (Figure 4.07 and Figure 4.08). The covariates used are the same covariates as were used in Chapter 4 on the distraction vOICe task. The covariate procedure is detailed in Chapter 4 methods, and the individual covariate details are in the Chapter 4 results. The slope covariate showed a medial frontal area correlated with improvement at localization, and the intercept covariate showed that BA 19 ($p < 0.05$, small volume corrected) correlated with initial localization performance (Table 5.3). The visual activation result for the localization intercept covariate is interesting, and verifies that the visual activation in the vOICe localization task is likely due to vOICe interpretation. The correlation with initial performance (intercept) rather than improvement (slope) may be due to the fact that the localization task is quite intuitive from the beginning, and therefore the intuitive existing crossmodal correspondences play a strong role in the visual activation from vOICe.

A



B

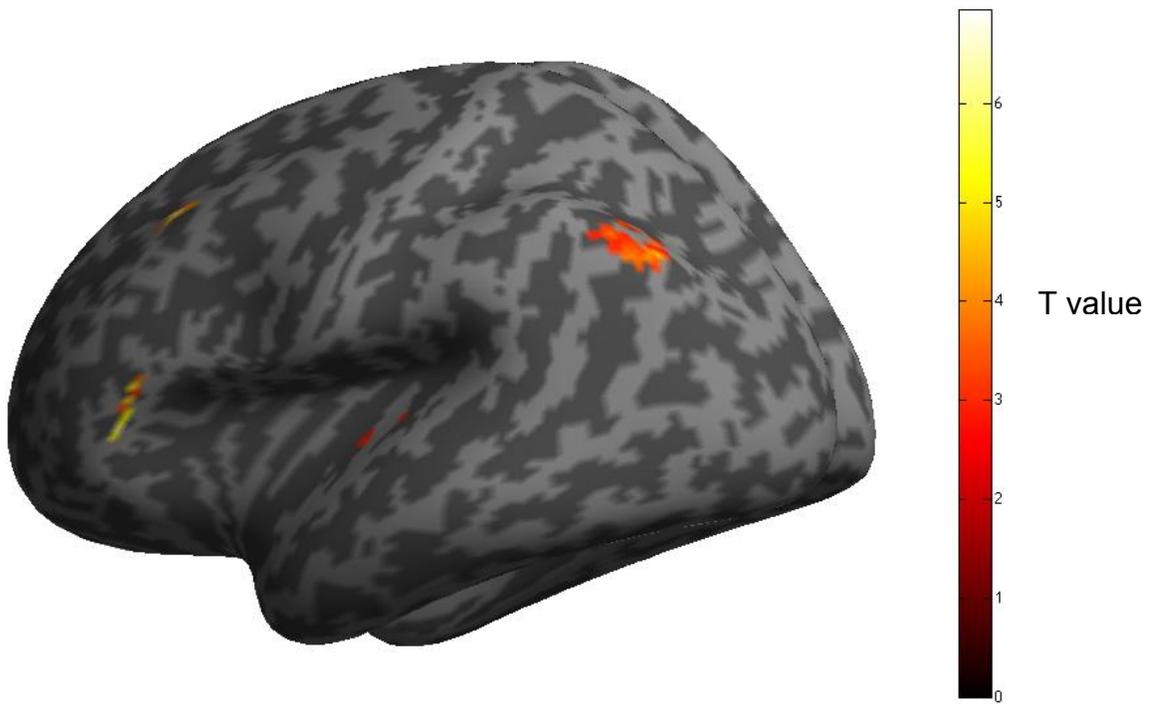


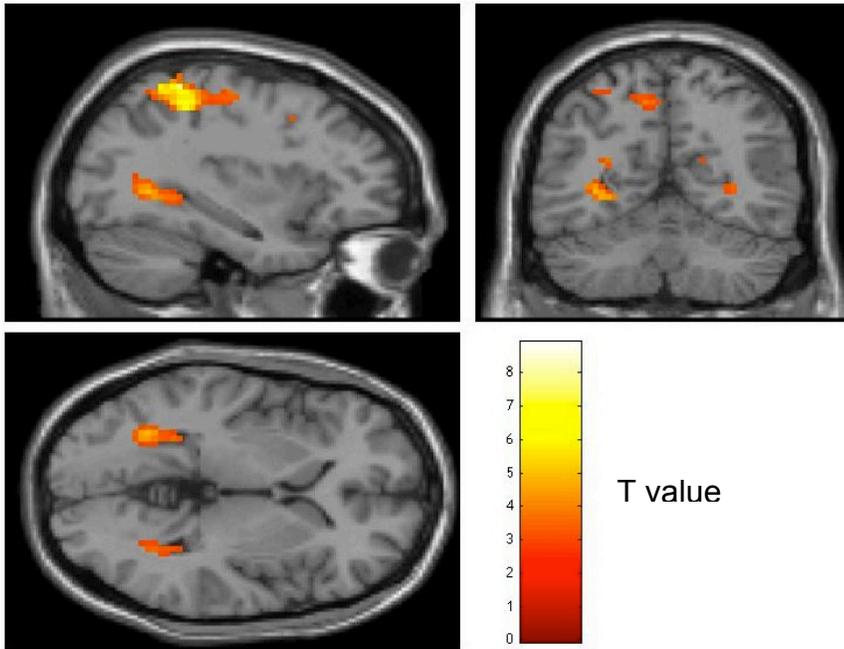
Figure 5.5. fMRI imaging results: vOICE dot [Right – Left location] post-scan in sighted participants. The neural imaging result is displayed for the post-vOICE-training right dot in contrast to the post-vOICE-training left dot both presented in vOICE with sighted participants ($N = 10$). Imaging data presented is $p < 0.009$ uncorrected and clusters of 10 voxels or more; further correction for multiple comparisons is shown in Table 5.1. Methods for fMRI data display are in the Chapter 4 methods.

Sighted Participants ($N = 10$)						
Region	BA	Side	x	y	z	p_{uncorr}
<i>vOICe Dot Post [Right – Left]</i>						
Superior Frontal Gyrus	8	L	-30	29	46	0.000
Inferior Frontal Gyrus	45	L	-48	26	1	0.000
Precuneus	19	L	-42	-73	43	0.001
<i>- small volume-corrected peak</i>						0.043*
Angular Gyrus	39	L	-42	-61	37	0.001
<i>- small volume-corrected peak</i>						0.056*
Medial Frontal Gyrus	10	R	9	47	10	0.001
Superior Temporal Gyrus	22	L	-48	-7	1	0.004
<i>- small volume-corrected peak</i>						0.112*
Insula	13	L	-42	-19	1	0.005
<i>vOICe Dot Post [Left – Right]</i>						
Cingulate Gyrus	32	L	-15	20	31	0.000
Cingulate Gyrus	24	L	-9	-4	28	0.000
Cingulate Gyrus	24	L	-15	-1	34	0.001

Table 5.1. fMRI imaging results: vOICe dot post-scan sighted participants. Imaging results for sighted participants when comparing the post-training left dot and the post-training right dot in vOICe ($N = 10$). All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold. The small volume correction was for a sphere of 10 millimeters radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed

on the talarach client for nearest grey matter. Any clusters without nearest grey matter within ± 5 mm are not included.

A



B

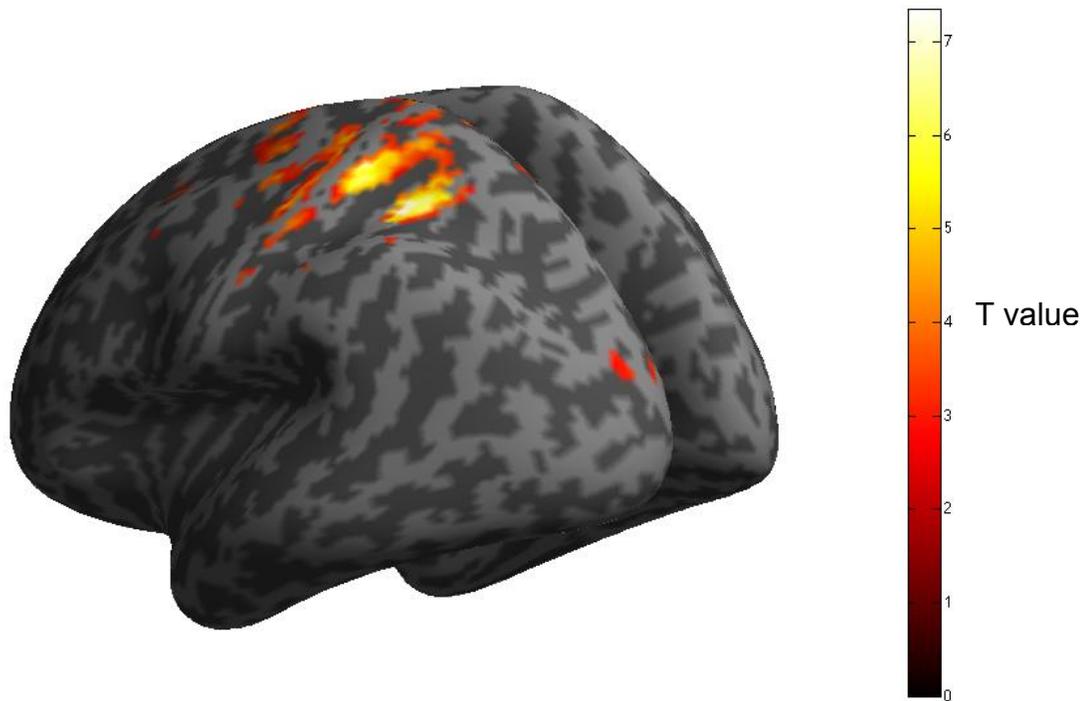
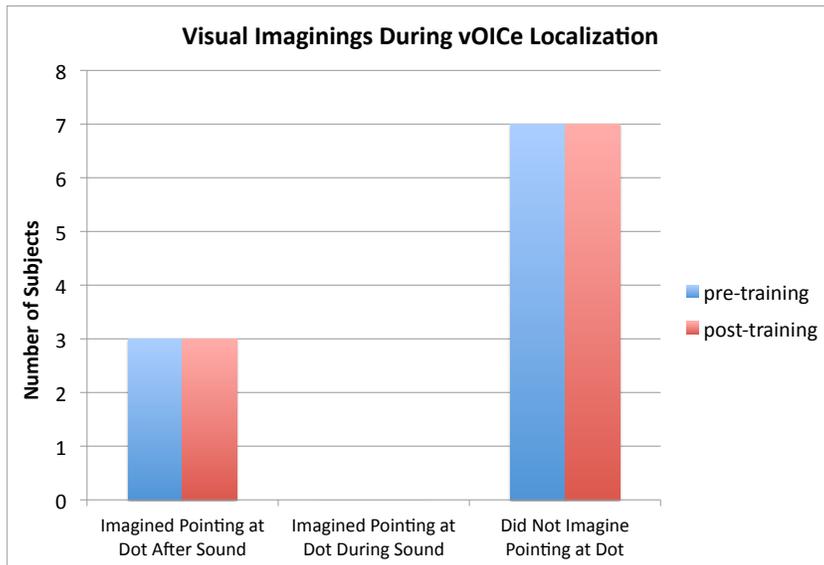


Figure 5.6. fMRI imaging results: Vision dot [Right – Left location] post-scan in sighted participants. The neural imaging result is displayed for the post-vOICE-training right dot in contrast to the post-vOICE-training left dot, both presented in vision with sighted participants ($N = 10$). Imaging data presented is $p < 0.009$ uncorrected and clusters of 10 voxels or more; further correction for multiple comparisons is shown in Table 5.2. Methods for fMRI data display are in the Chapter 4 methods.

Sighted Participants (N = 10)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>Vision Dot Post [Right – Left]</i>						
Inferior Parietal Lobule	40	L	-36	-43	52	0.000
Postcentral Gyrus	5	L	-36	-43	61	0.000
Superior Parietal Lobule	7	L	-33	-52	61	0.000
Lingual Gyrus	19	L	-33	-58	1	0.000
<i>- small volume-corrected peak</i>						0.016*
Parahippocampal Gyrus	19	L	-30	-46	-2	0.001
<i>- small volume-corrected peak</i>						0.029*
Angular Gyrus	39	R	48	-70	28	0.002
<i>- small volume-corrected peak</i>						0.039*
Posterior Cingulate	31	L	-27	-58	19	0.003
Cuneus	18	L	-18	-82	22	0.003
Cuneus	18	L	-15	-91	16	0.005
Middle Frontal Gyrus	6	L	-27	11	43	0.003
Cingulate Gyrus	24	R	9	2	46	0.003
Parahippocampal Gyrus	19	R	33	-55	1	0.004
Cingulate Gyrus	31	R	21	-46	22	0.005
<i>Vision Dot Post [Left – Right]</i>						
No significant activation						

Table 5.2. fMRI imaging results: Vision dot [Right – Left location] post-scan sighted participants. Imaging results for sighted participants when comparing the post-training left dot and the post-training right dot in vision ($N = 10$). All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold. The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within ± 5 mm are not included.

A



B

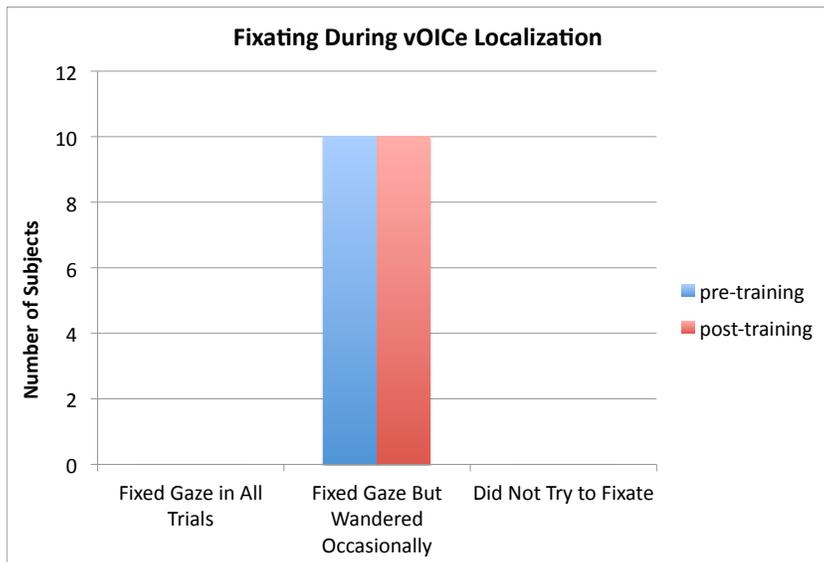


Figure 5.7. Post-experiment questionnaire results. Following the post-training fMRI scan, all participants filled out a questionnaire (Appendix C). This figure plots the responses to select questions in that questionnaire for the 10 sighted participants.

Sighted Participants ($N = 10$)						
Region	BA	Side	x	y	z	p_{uncorr}
vOICe [Right – Left] Post Localization Slope Covariate						
Medial Frontal Gyrus	6	L	–3	32	37	0.000
vOICe [Right – Left] Post Localization Intercept Covariate						
Parahippocampal Gyrus	19	L	–30	–49	1	0.001
- <i>small volume-corrected peak</i>						0.035*
Culmen, Cerebellum		L	–6	–58	1	0.003

Table 5.3. fMRI covariate data: vOICe dot [Right – Left location] post-scan sighted participants. Two covariates for vOICe dot [Right – Left location] post-scan are displayed in this table, both based on vOICe training performance. Details on the processing of covariates is in the methods section and the results section of Chapter 4 (same covariates as last two in Table 4.4). The visual neural activation shown for the vOICe dot [Right – Left location] post-scan correlates with the performance metric, indicating that the covariate may have played a role in generating the neural activation listed. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included. The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected.

Visually Impaired Neural Imaging Results

Four visually-impaired individuals performed the vOICe localization experiment: a severe low-vision participant, a late blind participant, and two congenitally blind participants (details on participants is in Chapter 4 methods, p. 130). The severe low-vision participant did not have any neural activation to the localization contrasts used (*i.e.* vOICe dot post [Right – Left location] or vOICe dot post [Left – Right location]). The results for the late blind participant and the congenitally blind participants are presented in Figure 5.8 and Table 5.4. The late blind participant ($N = 1$) had visual activation (BA 19, and 18) for the vOICe dot Post [Left – Right] in the right and left hemisphere (no activation for vOICe dot Post [Right – Left]). This is a bilateral mapping from visual space (left dot) to neural activation (right and left hemisphere). The first congenitally blind participant, WB, had quite different results: visual activation (BA 17 and 18) for vOICe dot Post [Right – Left] in the right hemisphere. The congenitally blind participant WB, therefore has an ipsilateral mapping from visual space (right dot) to visual neural activation (right hemisphere). The second congenitally blind participant, SB, had visual activation (BA 19) for the vOICe dot Post [Left – Right] in the left and right hemisphere. Therefore, SB had a bilateral mapping of vOICe crossmodal plasticity. The sighted participants' ($N = 10$) results reported earlier for vOICe dot localization post-training (Table 5.1) had a contralateral mapping from visual space to neural activation. Therefore, interestingly, the sighted participants had a contralateral mapping, the late blind participant had a bilateral mapping, and the congenitally blind participants had an ipsilateral and bilateral mapping. Although there are too few blind participants ($N = 2$) to make strong conclusions, there is a trend for visual experience to be associated with a

contralateral mapping via vOICE, and lack of visual experience to be associated with an ipsilateral or bilateral mapping via vOICE.

A

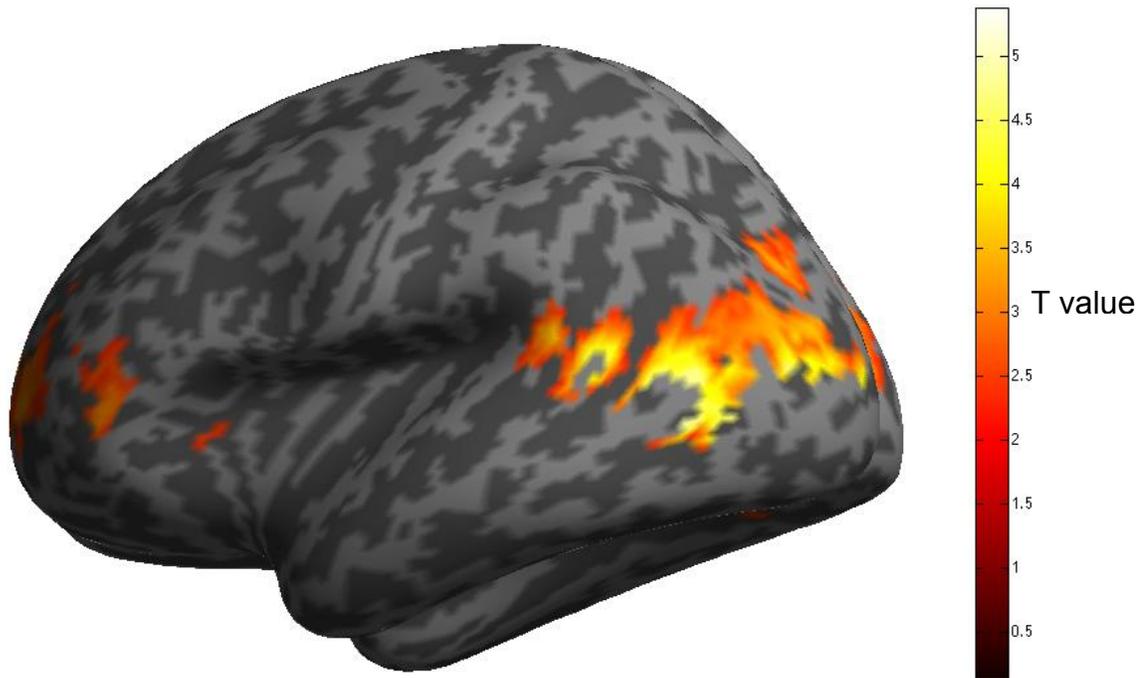


Figure 5.8. fMRI imaging results: vOICE dot [Right – Left location] post-scan congenitally blind participant ($N = 1$). The neural imaging result is displayed for the post-vOICE-training right dot in contrast to the post-vOICE-training left dot, both presented in vOICE with a blind participant ($N = 1$). Imaging data presented is $p < 0.009$ uncorrected and clusters of 10 voxels or more; further correction for multiple comparisons is shown in Table 5.4. Methods for fMRI data display are in the Chapter 4 methods. A severe low-vision participant ($N = 1$) also performed this experiment, but had no significant neural activation for this contrast.

A

Late Blind Participants (N = 1) (RD)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>vOICe Dot Post [Right – Left]</i>						
No Activation						
<i>vOICe Dot Post [Left – Right]</i>						
Fusiform Gyrus	37	R	42	-55	-8	0.000
- small volume-corrected peak						0.000*
Clastrum			36	-22	-2	0.000
Fusiform Gyrus	19	R	42	-73	-11	0.000
- small volume-corrected peak						0.012*
Temporal Lobe	37	L	-42	-46	-8	0.000
Culmen		L	-18	-58	-8	0.000
Culmen		L	-21	-49	-11	0.000
Cuneus	18	R	15	-67	16	0.000
- small volume-corrected peak						0.003*
Posterior Cingulate	30	R	15	-52	13	0.000
Cuneus	18	R	12	-76	25	0.000
Middle Temporal Gyrus	39	R	51	-76	25	0.000
- small volume-corrected peak						0.007*
Middle Temporal Gyrus	39	R	57	-67	25	0.000
Middle Temporal Gyrus	39	R	60	-64	13	0.003
Middle Occipital Gyrus	18	L	-24	-82	-8	0.000
Thalamus		L	-3	-7	10	0.000
Lentiform Nucleus		L	-18	2	10	0.002

B

Congenitally Blind Participants (N = 1) (WB)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>vOICe Dot Post [Right – Left]</i>						
Lingual Gyrus	17	R	18	–88	4	0.000
<i>- small volume-corrected peak</i>						0.000*
Lingual Gyrus	17	R	6	–85	4	0.000
Middle Occipital Gyrus	18	R	30	–79	1	0.000
<i>- small volume-corrected peak</i>						0.000*
Medial Frontal Gyrus	10	R	15	50	7	0.000
Medial Frontal Gyrus	10	L	–9	47	4	0.000
Medial Frontal Gyrus	10	R	18	68	4	0.000
<i>vOICe Dot Post [Left – Right]</i>						
Thalamus		L	–21	–16	10	0.000
Thalamus, Medial Dorsal Nucleus		L	–6	–13	4	0.001
Thalamus		R	24	–25	10	0.000

C

Congenitally Blind Participants (N = 1) (SB)						
Region	BA	Side	x	y	z	p_{uncorr}
<i>vOICe Dot Post [Right – Left]</i>						
No Activation						
<i>vOICe Dot Post [Left – Right]</i>						
Postcentral Gyrus	3	R	42	-25	61	0.000
Superior Frontal Gyrus	6	R	9	-16	61	0.000
Medial Frontal Gyrus	6	R	12	-28	58	0.000
Supramarginal Gyrus	40	R	60	-58	34	0.000
<i>- small volume-corrected peak</i>						0.000*
Inferior Parietal Lobule	40	R	66	-40	40	0.000
Middle Temporal Gyrus	39	R	54	-70	28	0.000
<i>- small volume-corrected peak</i>						0.002*
Precuneus	19	R	3	-88	46	0.000
<i>- small volume-corrected peak</i>						0.009*
Cuneus	19	R	12	-76	34	0.002
Cuneus	18	R	18	-79	25	0.004
<i>- small volume-corrected peak</i>						0.058*
Superior Temporal Gyrus	42	L	-60	-25	13	0.001
Postcentral Gyrus	43	L	-54	-7	19	0.001
Transverse Temporal Gyrus	42	L	-60	-16	16	0.001
Inferior Parietal Lobule	40	L	-63	-46	43	0.001
Precuneus	19	L	-42	-82	40	0.001
Angular Gyrus	39	L	-54	-70	40	0.001
Precuneus	19	L	-45	-76	46	0.001
Superior Temporal Gyrus	41	L	-42	-40	7	0.003

Congenitally Blind Participants ($N = 1$) (SB)						
Region	BA	Side	x	y	z	p_{uncorr}
Superior Temporal Gyrus	41	L	-42	-40	7	0.003
Superior Temporal Gyrus	41	L	-45	-31	10	0.004
Postcentral Gyrus	2	L	-39	-28	34	0.003
Cingulate Gyrus	31	R	24	-49	31	0.005

Table 5.4. fMRI data: vOICE dot [Right – Left location] post-scan blind participants. Select imaging results for blind participants when comparing the post-training left dot and the post-training right dot in vOICE ($N = 1$ for late blind and $\underline{N} = 1$ for congenitally blind). For the late blind participant, only the top 15 clusters of activation are presented in Panel A; a full list is in Appendix D, Table C. All other participant sub-tables contain a full list of neural activation. All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold. The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talaraiich client for nearest grey matter. Any clusters without nearest grey matter within ± 5 mm are not included. A severe low-vision participant ($N = 1$) also performed this experiment, but had no significant neural activation for this contrast.

Discussion

This chapter focused on testing the topographical mapping of sensory substitution perception. The main experiment used the localization of a dot in the left or right visual field encoded into sound to determine whether it is contralaterally mapped like vision (*i.e.*, left dot activates right visual regions). This main experiment showed that after the vOICE training, visual region V3 was activated contralaterally for the vOICE right dot (*i.e.*, left hemisphere), but that this mapping was not present for the vOICE left dot. In fact, no visual activation occurred more strongly for the vOICE left dot when compared to the right dot encoded with vOICE. Interestingly, after training, a similar pattern occurred in the visual control (images of the dots were used rather than the sounds). The activation in left visual cortex may be due to dominance of the right side (hand dominance *etc.*) for most individuals, therefore making the left hemisphere more dominant. This localization task, which closely matches the localize, reach, and touch task in the office (detailed in Chapter 4 methods, Figure 4.07), may be engaging the left lateralized mirror system, which ranges from pre-motor, temporal and parietal regions (Ricciardi et al., 2009). In this case, the participants' motor mirror system would be mirroring a remembrance of performing the reach and touch task that they just performed in the lab outside the fMRI scanner 10 to 30 minutes before. While the topographic mapping does not entirely mimic normal vision's contralateral mapping, the visual control and vOICE sound were quite similar in scans following training, and it is possible that crossmodal plasticity from training caused this similarity between vision and vOICE mapping.

The post-training questionnaire was used to determine whether visualization played an important role in visual activation from vOICE sounds. The questionnaire showed that participants either visualized following the sound, or not at all. The scans used for analysis were during the vOICE sound; therefore, the contrasts did not capture any of the visualization by participants. This self-reporting evidence indicates that the visual activation from the vOICE task was not likely due to visualization by the participants. Further, a covariate analysis was used to determine whether any of the vOICE training performance correlated with the visual activation. The covariate for initial performance on vOICE (intercept) did correlate with visual activation in the vOICE task, indicating that the visual activation during the vOICE localization task was likely based on vOICE interpretation. This covariate is further evidence that it is the crossmodal interpretation of vOICE that originated the visual activation during the vOICE localization task.

The results from this chapter are not particularly definitive on the topographic mapping of vOICE sounds. The vOICE and vision mappings after vOICE training were similar but, unlike normal vision, strangely anisotropic. The fact that both vision and vOICE were asymmetrical may indicate that the localization task is too repetitive and simple, and therefore reducing visual activation. From another perspective, it could be the vOICE training itself that is impacting visual processing (as seen in Chapter 4). Just looking at the vOICE results, it is possible that more experience with vOICE is needed (than 5 hours of training) to generate a neurotypical contralateral mapping of space to visual activation. Perhaps an advanced user of sensory substitution would have a more solidified and consistent mapping generated from years of device use. In fact, the

echolocation spatial mapping of visual activation existed in an expert echolocator that had years of experience (Milne, et al., 2013). Therefore, these results likely indicate the beginning formation of a spatiotopic map of visual space with sensory substitution that is not yet fully complete.

Introduction

The broad themes of this thesis have ranged from crossmodal plasticity to automaticity (behavioral and neural), and rehabilitation of the blind population. Crossmodal plasticity is critical to the learning of any sensory substitution encoding, as sensory substitution inherently bridges across two modalities: the sense that receives the information, and that which interprets it. The automaticity of sensory substitution was studied both behaviorally (Chapter 3) and with neural imaging (Chapter 4). Automaticity of SS is critical to improving blind rehabilitation with sensory substitution, and the studies in this thesis will aid in the development of better training techniques and device encodings. Finally, blind rehabilitation has recurred as a theme throughout all of the thesis chapters, and is an important end application of this research.

Discussion***Crossmodal Plasticity***

Crossmodal plasticity is the foundation of all sensory substitution learning. Through crossmodal interactions and then plastic changes of those interactions, sensory substitution stimuli are interpreted visually, and action is generated. The type of plasticity, whether strengthening or weakening of existing neural connections or the generation of new neural connections, likely depends on the task, duration of training, and visual deprivation of the participant (*i.e.*, blind or sighted).

The experiments in this thesis all rely on plastic changes across the senses to generate improved performance at sensory substitution tasks. The results of these plastic changes are measured behaviorally in Chapters 2 and 3, and with neural imaging (fMRI) in Chapters 4 and 5. In Chapter 2, the constancy processing of SS stimuli (after training) is likely mediated by visual neural regions that are activated by crossmodal plastic changes. Chapter 3 studied the underlying crossmodal mappings that are used in the interpretation of SS by naïve and trained users. Some intrinsic correspondence/mapping seemed to exist, mediating A-V matching performance in the trained as well as in the naïve participants. These crossmodal neural connections generating the crossmodal mappings are potentially strengthened via SS training to generate relevant improvements in performance. In Chapter 4 and 5, crossmodal plasticity is measured explicitly with fMRI scans before and after vOICE training. Chapter 4 determines whether the crossmodal plasticity can be activated automatically (*i.e.*, without attention) after training on an SS device. This was confirmed via a mental counting task that distracted attention while a vOICE encoding of white noise was played. In Chapter 5, the mapping from visual space through SS to visual activation is measured to determine whether the crossmodal plasticity is topographically mapped. Both Chapter 4 and 5 serve to better understand crossmodal plasticity with sensory substitution by testing its automaticity and spatial mapping.

Intrinsic Crossmodal Mappings

Intrinsic mappings across the senses (such as vision and audition) were shown to be important to sensory substitution interpretation in Chapter 3. Chapter 3 studied whether any vOICE sounds could be intuitive without any knowledge of vOICE by using

the crossmodal mappings (such as matching a high pitch with a high spatial location) that participants already had. Surprisingly, the naïve could interpret vOICe sounds, and could do so automatically (independent of attentional load). Given this result in Chapter 3, it is likely that crossmodal mappings play a key role in the sensory substitution learning in each of the other chapters, and may even underlie a part of the visual activation in response to vOICe sounds.

Automaticity

Automaticity was the key concept in Chapters 3 and 4 to study the assumed cognitive (top-down) nature of sensory substitution interpretation. In general, SS is limited in its commercial prospects due to the long training time and the heavy cognitive burden of interpretation. Therefore, we have studied in this thesis ways to make SS more automatic. In Chapter 3, we investigated crossmodal mappings (such as matching a high pitch to high spatial position) to determine whether images and encodings with crossmodal mappings can be easy or automatic to interpret. We found that these intuitive and existing mappings made vOICe interpretation attention-load insensitive (*i.e.*, independent of attention) even in entirely naïve users. In Chapter 4, we investigated if the crossmodal plasticity generated by using SS can also be automatic. This fMRI experiment used a distraction task to test for attention-load sensitivity. The results showed that visual activation generated by crossmodal plasticity was not dependent on attention.

The study of automaticity and sensory substitution is quite novel. Because SS is assumed to be top-down and cognitively intensive (or rather, no researchers had paid attention to this dimension of top-down attentive *vs.* automatic), no studies have

investigated whether there is an element of SS that might be intuitive or processed automatically. The study of intrinsic crossmodal mappings and their role in making SS interpretation automatic (in Chapter 3) is the first step in highlighting the automatic elements of SS and expanding their role in SS. The study of the automaticity of crossmodal plasticity following training with SS (Chapter 4) is a novel indication that plasticity engendered by SS usage can be automatic (*i.e.*, not require attention). These investigations may allow for improvements in training to tap into this automatic crossmodal plasticity and make SS easier to use.

This thesis provided two critical results on the automaticity of SS that should be emphasized. The first result, from Chapter 3, is that if existing crossmodal connections and mappings are optimally used in stimuli and encodings, then SS can be automatically interpreted. The second finding, in Chapter 4, indicated that crossmodal plastic changes engendered by training can be automatically activated independent of attentional demands. Combined, these results show that sensory substitution may have hope of becoming a more easily interpreted device, and consequently aid a wider blind population.

Blind Rehabilitation

Improving the capabilities of the blind is a major goal of sensory substitution as well as the research in this thesis. The blind are a large disabled population within the United States and around the world. An inexpensive and useful aid for the blind could help not only individuals in advanced countries, but also those throughout impoverished nations. Sensory substitution has the potential to be this device.

The research in this thesis aims to improve SS devices with psychophysical as well as neural imaging studies. In Chapter 2, the functional use of SS to externalize vOICE stimuli via shape and rotation constancy is an important step toward the processing of objects in space and in the correct proportion and orientation. Chapter 3 focuses on making SS easier to interpret by using intrinsic crossmodal mappings that users already have. More ease of use could make sensory substitution a better aid to the blind and therefore more widely utilized. Moreover, the results indicate that vision-like perception (in the sense of being effortless) can be accomplished via training potentially more easily than previously believed. In addition, Chapter 4 and 5 investigate the neural processing of SS, the results of which could be used not only to understand the neural mechanisms of multisensory plasticity, but also to optimize device training to generate more crossmodal plasticity from SS use. Greater crossmodal plasticity would improve device performance, and thereby enhance rehabilitation. Overall, the behavioral studies in Chapter 2 and 3 directly test methods to improve blind rehabilitation with SS devices with promising results, and the neural imaging in Chapters 4 and 5 use enhanced understanding of neural processing as tools to improve SS device usage. Not only that, a part of the results further confirmed the attentionless, automatic nature of the perceptual interpretation after SS training. Therefore, the results in this thesis are important steps toward making SS devices more intuitive and utilizing the potential of crossmodal plasticity to improve device interpretation.

Interaction of Thesis Themes

The roles of the thesis themes (detailed above) as tools, experiments, and end goals are spatially laid out in Figure 6.1. The major neural processing capabilities have

been used as tools in this thesis, and include: Crossmodal plasticity and sensory motor learning, which were both used to train blind and sighted individuals to use the vOICe and to engender improvement during that training. The two major end goals from the experiments in this thesis are the rehabilitation of the blind and the advancement of neuroscientific understanding of multisensory mapping and plasticity, both of which were furthered in the execution of the thesis experiments. The vertical y -axis of Figure 6.1 shows that several chapters of this thesis are more basic-science-themed (the end aims are to advance the scientific understanding, rather than a material or physical goal). In contrast, other experiments are of a more applied-science nature, and strive to develop a device to aid the blind. Of course, the end goals have a moderate overlap across chapters, thereby generating the cross arrows.

An alternative method of visualizing the themes in this thesis is as a pyramid (Figure 6.2). The pyramids base blocks consist of the crossmodal plasticity and sensory motor learning, which then support two additional blocks: The automaticity of learning block, and then the blind rehabilitation block. With pyramid height corresponding to vOICe learning, each of the building blocks increases in vOICe learning, and is supported by the blocks beneath them. This visual analogy makes it clear why greater training techniques to enhance sensory motor learning and crossmodal plasticity are critical to the success of sensory substitution as an aid for the blind. If either of the foundation stones crumples, blind rehabilitation with sensory substitution will not succeed.

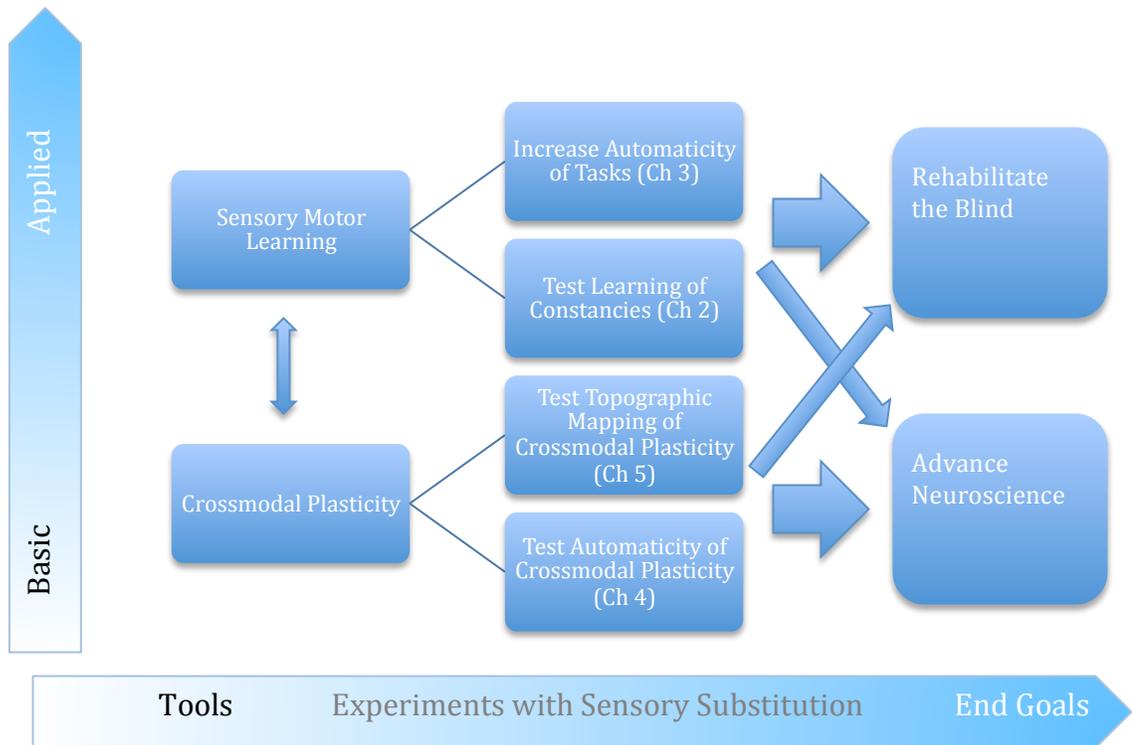


Figure 6.1. Concept web for thesis. This diagram spatially lays out the concepts developed in the thesis, and maps out several interesting inter-connections among concepts. In particular, it maps out the progress from tools to experiments to scientific goals for the thesis. It also shows the range from basic science to more applied science, and various cross-connections among the two.

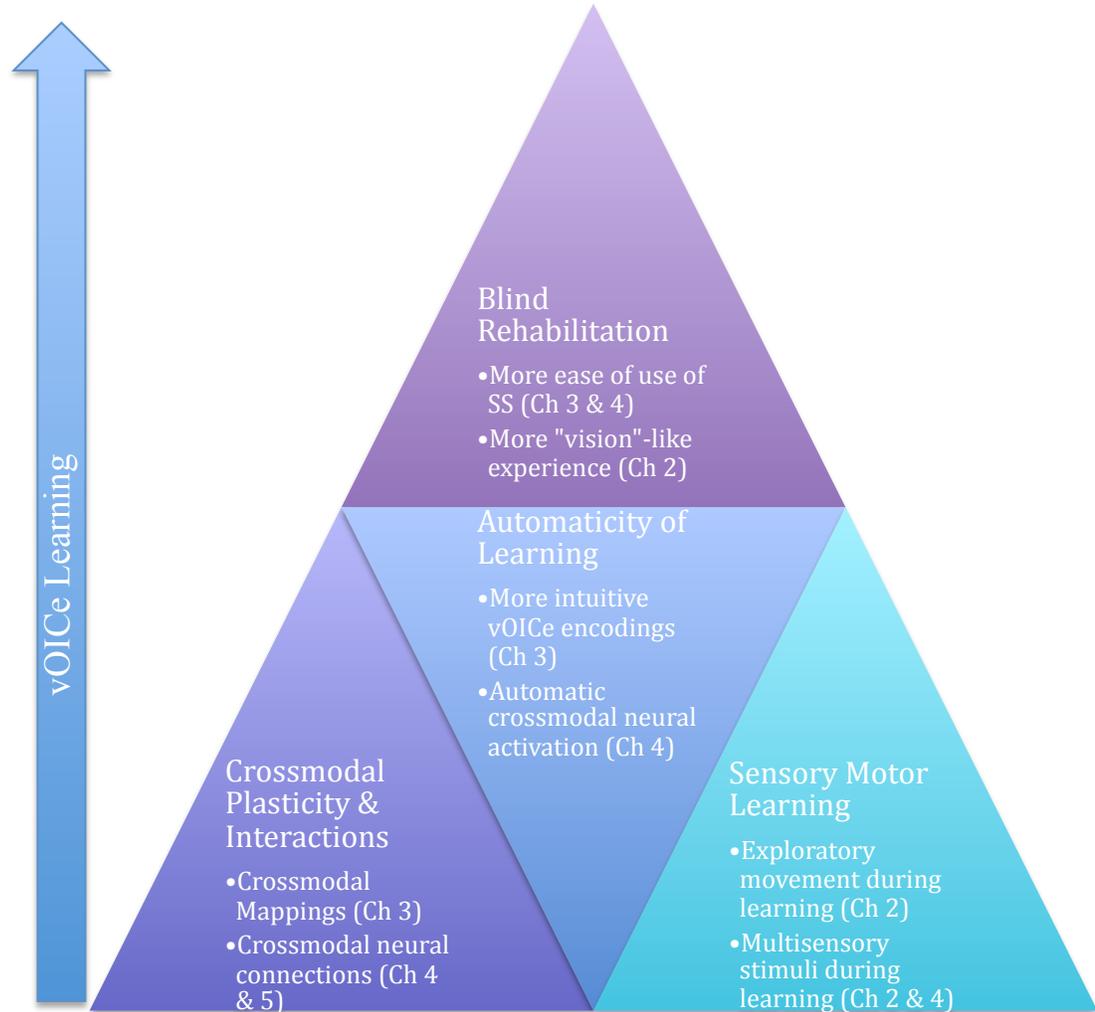


Figure 6.2. Layout of thesis themes. An alternative layout of thesis themes shows the crossmodal plasticity and sensory motor learning at the base of the pyramid, supporting the automaticity of perceptual processing and the rehabilitation of the blind. Each of the pyramid blocks has references to the chapters that relate strongly to those themes.

Research Next Steps

Research is a continuous process of discovery, and the studies in this thesis are just one step in a march toward understanding the brain. Therefore, there are several experiments and studies following on the work in this thesis that will continue to add to neuroscience. A few of these potential experiments are highlighted below.

Perceptual Constancy

Chapter 2 focused on the learning of constancies with the vOICe device; in particular, length constancy and shape constancy were learned by sighted and blind participants. Additional perceptual constancies would also be interesting to test with the vOICe device, such as size constancy (objects appear the same size independent of distance), which is valuable to monocular depth perception, or brightness constancy (objects appear the same brightness independent of lighting conditions), which is valuable to recognition and localization capabilities. Further, we tested constancies in a simplified lab setting; training and testing the use of constancies in daily-life tasks would be an important step toward full visual perception and capabilities. Such daily-life tasks may include recognizing and picking up an object on a table independent of object orientation (shape constancy) or lighting conditions (brightness constancy).

Neural Correlates of Intrinsic Crossmodal Mapping

In Chapter 3, it was shown that crossmodal correspondences generate the intuitiveness of different stimuli encoded by SS. This chapter used several behavioral psychophysical tests to determine the role of crossmodal mappings in sensory substitution interpretation, and the automaticity of interpreting crossmodal mapping-rich SS sounds. An interesting follow-up experiment would be to study the neural correlates

of the interpretation of SS based on intrinsic crossmodal mapping. In particular, it would be interesting if intuitive sounds that are crossmodal mapping-rich also have more visual activation (via crossmodal interactions) than SS sounds that are crossmodal mapping-deficient. This correlation between crossmodal mapping intuitiveness and visual activation (due to crossmodal interactions/plasticity) would indicate the neural processing behind the use of crossmodal mappings to interpret SS effortlessly.

Correlation with Other Multisensory Effects/Tasks

Another experiment using the premise of Chapter 3 (*i.e.*, crossmodal interactions impacting SS interpretation) would study whether participants that have strong crossmodal interactions also find SS more intuitive and easy to learn. Tests of crossmodal interactions could include bouncing *vs.* streaming effect, the double flash illusion, or the McGurk Effect. There is also a range of SS tests that could be used for this experiment including localization, recognition, and depth perception. The more similar the crossmodal interaction and SS task, the more likely that they will use similar multimodal pathways and therefore be correlated. Therefore, the bouncing *vs.* streaming effect and movement evaluations of speed and direction in SS would be more likely to be correlated than bouncing *vs.* streaming and object recognition. This line of research, if further applied to the blind population (V-T mapping), may eventually provide us with a simple diagnostic test of suitability of SS to a particular individual.

Testing Effects of SS Training by Multisensory Illusions

In the same direction, SS training and the resulting crossmodal plasticity may impact the strength of existing crossmodal interactions. In this experiment, the strength of a crossmodal illusion could be tested before and after training on sensory substitution.

As with the comparison above, the more similar the SS training and the crossmodal interaction, the more likely that SS training will impact the strength of the crossmodal interaction. It is also more likely that crossmodal interaction strengthening will be detected if it is tested as soon after training as possible.

Suppression of Visual Cortical Processing by SS Training

In Chapter 4, fMRI imaging was used to test whether crossmodal plasticity from vOICe training was automatic (or engaged without attention). As a part of this chapter, it was found that visual activation due to a vision white noise pause detection task was suppressed following training relative to before training in sighted individuals. It would be interesting to determine whether this suppression effect only occurs with white noise images, or if it also occurs with other images and/or visual tasks. Further, does the visual suppression correlate with the amount of crossmodal plasticity in each individual? Deeper investigation of this suppression phenomenon may lead to interesting conclusions on the competition between visual and crossmodal processing in the brain.

Conclusion

This thesis has used psychophysics and neural imaging to study crossmodal plasticity and improve blind rehabilitation with sensory substitution. The results contribute to the understanding of neural changes, and add new crossmodal methods to improving sensory substitution for blind rehabilitation. New experiments based on the results in this thesis are plentiful, including new studies on crossmodal mappings and SS crossmodal plasticity. New research will hopefully build upon this thesis's results to construct a better understanding of the brain, and through that understanding aid populations recover from neural deficits.

SUPPLEMENTARY DATA FOR CHAPTER 3

Figure A-C: This figure contains the task-performance matching images to vOICe sounds of naïve and trained participants for all image sets tested in Chapter 3. It also contains the pvalue threshold markers for the comparison to chance of naïve and trained data, as well as the naïve to trained comparison. The blue and red stars indicate that a given image set is significantly different from chance ($p < 0.05$) for the naïve and trained individuals, respectively. The purple stars indicate that the naïve and trained performance were significantly different from each other ($p < 0.05$).

Figure A

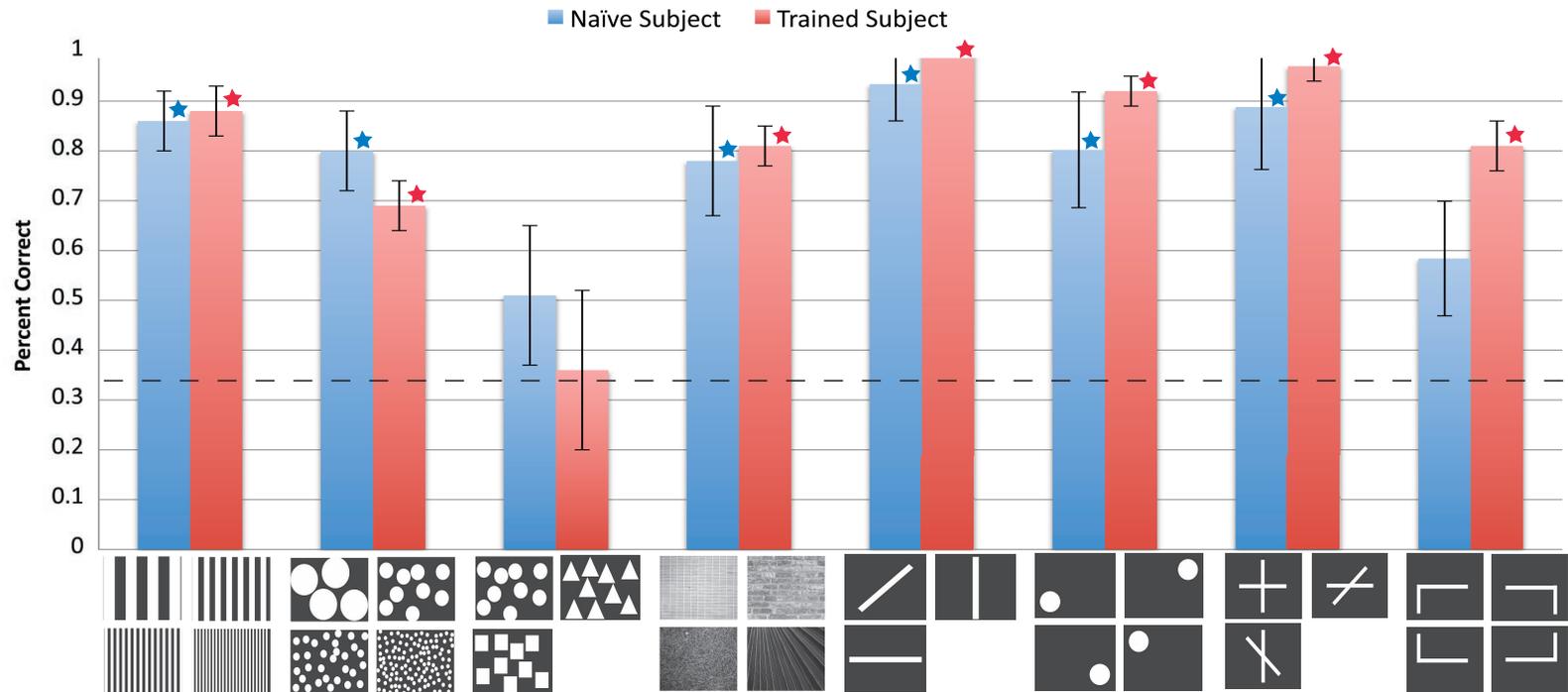


Figure B

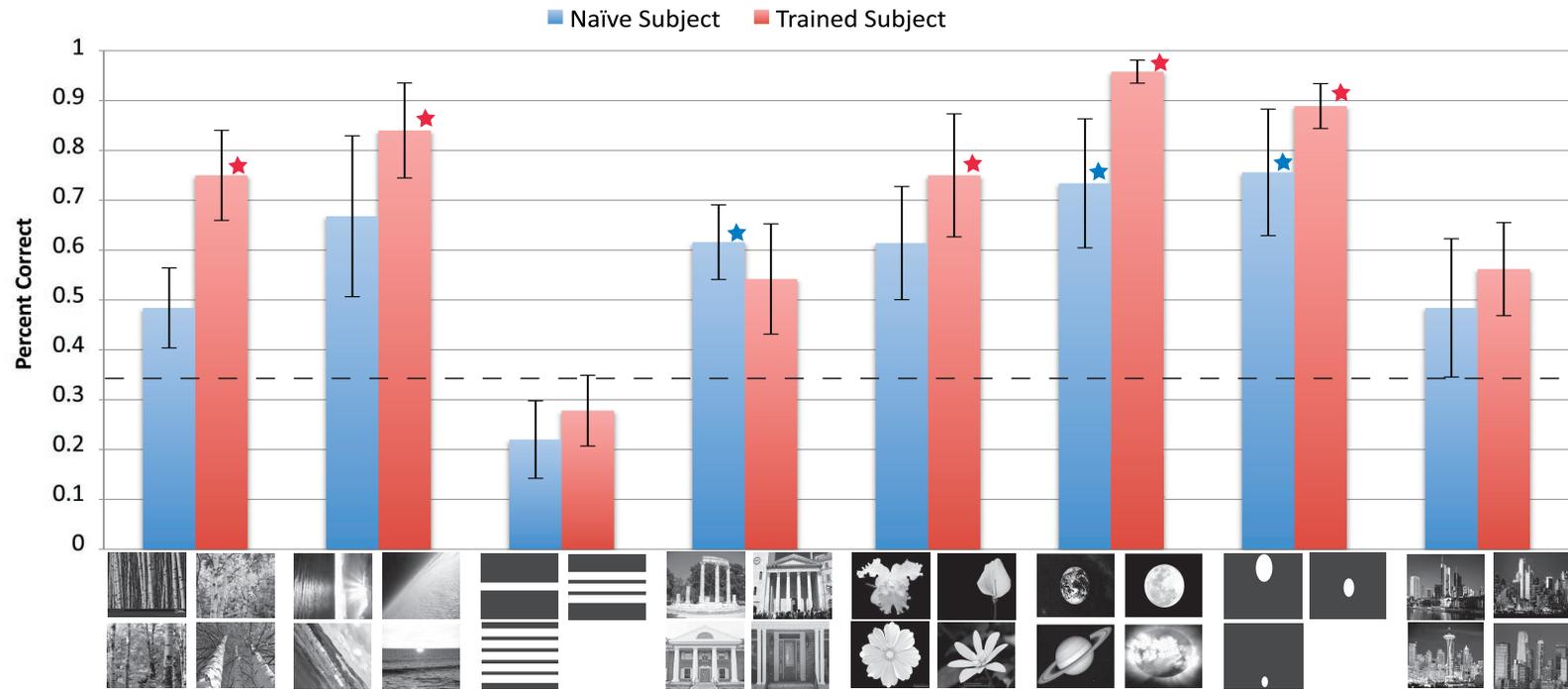
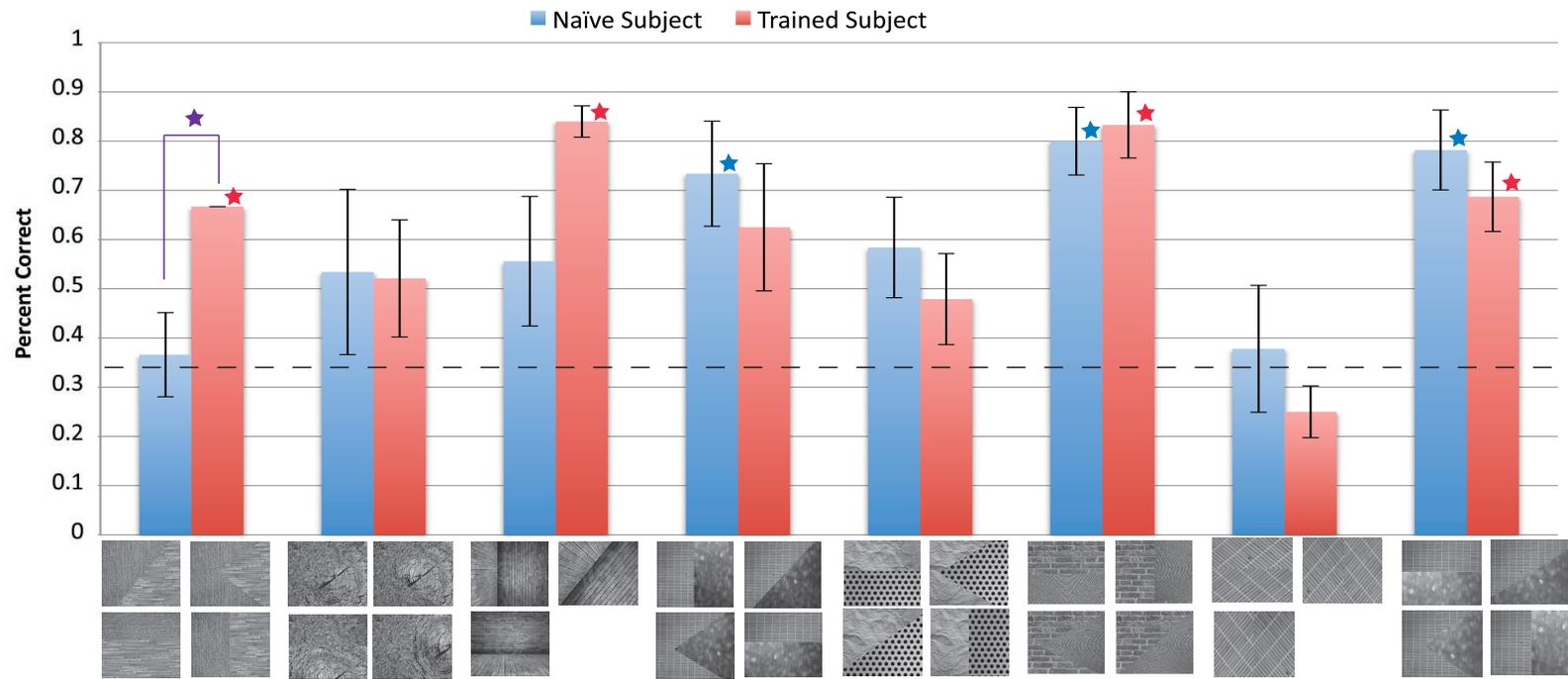


Figure C



APPENDIX B

vOICE TRAINING PROCEDURES

This appendix includes the detailed training instructions for the fMRI vOICE experiment (Chapter 4 and 5) in part 1, and the vOICE behavioral experiments (Chapter 2 and 3) in part 2. The instructions were drafted before and during training as a guide to the experimenter on the training procedure. Additional detail was added following the experiments to clarify the training procedures.

Appendix B Part 1

vOICe fMRI Localization Experiment Training Instructions

Session 1 (1 hour)

- Training Assessment (always perform assessment first)
 - o 10 trials of reaching for a white circle in one of five locations on a black felt covered board while sitting at a black-felt-covered table (positions randomized in MATLAB)
 - o Record accuracy of reaching before physically correcting the participant's reach to the center of the circle.

- Training Tasks:
 - o Locating, centering in the field of view and reaching for large circle on the black felt board (give feedback on the accuracy of centering before the participant reaches).
 - o Differentiating between configurations of white blocks and shapes on the black felt board (L from a backwards L, from a 7 and a backwards 7, and a circle from a square, from a rectangle).

Session 2 (1 hour)

- Training Assessment (always perform assessment first)
 - o 10 trials of reaching for a white circle in one of five locations on a black-felt-covered board while sitting at a black-felt-covered table (positions randomized in MATLAB)

- Record accuracy of reaching before physically correcting the participant's reach to the center of the circle.
- Tasks:
 - Locating, centering in the field of view and reaching for large circle on the black felt board (give feedback on the accuracy of centering before the participant reaches).
 - Localize, walk to, and touch a large circle (5.5 inches in diameter) on a black felt wall. The participant must center the object, walk several steps, and then re-center the object in iterations until the participant is within reaching distance. The experimenter walks the participant through the first trial, and then in future trials, allows the participant to independently perform the task, only indicating when the participant is within reaching distance of the black felt wall. The circle can be placed on the center, left or right, and high, mid-level or low on the wall.

Session 3 (1 hour)

- Training Assessment (always perform assessment first)
 - 10 trials of reaching for a white circle in one of five locations on a black-felt-covered board while sitting at a black-felt-covered table (positions randomized in MATLAB)
 - Record accuracy of reaching before physically correcting the participant's reach to the center of the circle.
- Tasks:

- Localize, walk to, and touch a large circle (5.5 inches in diameter) on a black felt wall. The participant must center the object, walk several steps, and then re-center the object in iterations until the participant is within reaching distance. The circle can be placed on the center, left or right, and high, mid-level or low on the wall.
- Avoid a white chair obstacle on the way to localizing and reaching for a large circle on the black felt wall. The participant must locate the chair, avoid the chair without touching it, and then localize the white circle. The chair can be placed in front of the participant, or to the left or to the right of the participant.

Session 4 (1.5 hours)

- Training Assessment (always perform assessment first)
 - 10 trials of reaching for a white circle in one of five locations on a black-felt-covered board while sitting at a black-felt-covered table (positions randomized in MATLAB)
 - Record accuracy of reaching before physically correcting the participant's reach to the center of the circle.
- Tasks:
 - Avoid a white chair obstacle on the way to localizing and reaching for a large circle on the black felt wall. The participant must locate the chair, avoid the chair without touching it, and then localize the white circle. The

chair can be placed in front of the participant, or to the left or to the right of the participant.

- Differentiate five office objects (scissors, stapler, tape dispenser, tissue box, and envelope) at the black felt covered table and board. Participants are shown the objects with the vOICe device and then are asked to identify the objects when presented in random order (order generated by experimenter, not computer).
- Train for the fMRI Experiment: Perform the localization of a white dot on the left or right with 1. visual stimuli alone on computer, 2. simultaneous vision and auditory stimuli (*i.e.*, vOICe) on computer and then 3. just auditory stimuli (*i.e.*, vOICe) alone (this training bridges between the just auditory and just visual ends of the experiment).

Session 5 (0.5 hours)

- Training Assessment

- 10 trials of reaching for a white circle in one of five locations on a black-felt-covered board while sitting at a black-felt-covered table (positions randomized in MATLAB)
- Record accuracy of reaching before physically correcting the participant's reach to the center of the circle.

vOICe Behavioral Experiments Training Instructions

Note: Several different experiments were attempted in the pre- and post- training behavior sessions (session 0 and session 10), including the texture experiments (Chapter 3). The experiments listed in session 0 and session 10 are just examples of those tested.

Session 0 (1 hour) (Performed on iMac computer)

- Bouncing vs. Streaming Experiment
 - o File: BounceVStream.m
- Moving Dot Experiment: Left-to-Right vs. Right-to-Left Rate Estimation Task (use headphones on table next to iMac computer)
 - o File: vOICeVisIIIExptMovDot2AFCQuarter.m

Session 1 (1 hour)

- Assessments Tasks:
 - o Shape Constancy Test: 20 trials of participants assessing bar length (lengths 1-5) independent of angle. Perform task on vOICe, and then with normal vision. Note: Allow participants to see the line lengths vertical and horizontal with vOICe for each length before beginning the test (**allow head tilt**).
 - o Rotation Constancy Test: 15 trials of participants assessing bar angle (0, 90, 45, -45, 22, or -22 degrees relative to vertical) independent of head

tilt. Note: Allow participants to see each angle and tilt their head to left and right while viewing each angle before beginning the test.

- Localization Trials: 10 trials of localizing a white dot on a black felt board with the vOICe device (5 separate positions). Record accuracy of reaching. Also record accuracy of random reaching for 10 trials (without vision, eyes closed), and with vision 10 trials (eyes open).
- Training Tasks:
 - Centering a white circle on the black-felt-covered table
 - Recognition of simple objects (such as distinguishing a square, triangle, and circle)
 - Distinguishing an “L” from a backward L, an upside-down L, and backward and upside-down L (*i.e.*, a 7)

Session 2 (1 hour) through Session 7 (1 hour)

- Assessments Tasks:
 - Shape Constancy Test: 20 trials of participants assessing bar length (1-5) independent of angle. Perform task on vOICe, and then with normal vision. Note: Allow participants to see the line lengths vertical and horizontal for each length before beginning the test (**allow head tilt**).
 - Rotation Constancy Test: 15 trials of participants assessing bar angle (0, 90, 45, -45, 22, or -22 degrees relative to vertical) independent of head

tilt. Note: Allow participants to see each angle and tilt their head to left and right while viewing each angle before beginning the test.

- Localization Trials: 10 trials of localizing a white dot on a black felt board with the vOICe device (5 separate positions). Record accuracy of reaching. Also record accuracy of random reaching for 10 trials (without vision, eyes closed), and with vision 10 trials (eyes open).

- Training Tasks:

- Work on shape constancy: Estimate length for just 90-degree lines, and then estimate length for just 45-degree lines (do not train on 0 or -45 degree angles) (Note: The training angles were limited to two angles for each participant, although the angles used across participants may have varied).
- Work on rotation constancy: Estimate angles with the head only vertical, then estimate angles with head tilted to the left only, and estimate angles with head tilted to the right only.

Session 8 (1 hour) – Session 9 (1 hour)

- Assessments Tasks:

- Shape Constancy Test: 20 trials of participants assessing bar length (1-5) independent of angle. Perform task on vOICe, and then with normal vision. Note: Allow participants to see the line lengths vertical and

horizontal for each length before beginning the test (**do NOT allow head tilt**).

- Rotation Constancy Test: 15 trials of participants assessing bar angle (0, 90, 45, -45, 22, or -22 degrees relative to vertical) independent of head tilt. Note: Allow participants to see each angle and tilt their head to left and right while viewing each angle before beginning the test.
 - Localization Trials: 10 trials of localizing a white dot on a black felt board with the vOICe device (5 separate positions). Record accuracy of reaching. Also record accuracy of random reaching for 10 trials (without vision, eyes closed), and with vision 10 trials (eyes open).
- Training Tasks:
- Work on shape constancy: Estimate length for just 90-degree lines, and then estimate length for just 45-degree lines (do not train on 0 or -45 degree angles)
 - Work on rotation constancy: Estimate angles with the head only vertical, then estimate angles with head tilted to the left only, and estimate angles with head tilted to the right only.

Session 10 (1.5 hour) (performed on iMac computer)

- Bouncing vs. Streaming
 - File: BounceVStream.m

- Moving Dot Experiment: Left-to-Right vs. Right-to-Left Rate Estimation Task (use headphones on table next to iMac computer)
 - o File: vOICeVisIIIExptMovDot2AFCQuarter.m
- Texture Experiment: Texture Interface V3 part II, and Texture V2 part I and part II
 - o Files: TextureR3_partII.m (in Texture Interface V3), TextureR1_part1.m (in Texture V2), TextureR1_partII.m (in Texture V2)

APPENDIX C

POST-FMRI SCANNING QUESTIONNAIRE

All fMRI participants filled out a questionnaire following their final fMRI scanning session of the vOICe fMRI experiment. This questionnaire was used to better process the fMRI data, and to take into account factors such visualization.

vOICe fMRI Subject Questionnaire

Name: _____

Date: _____

Thank you for performing the vOICe fMRI experiment. Please try to answer the following questions to the best of your memory.

1. I responded to questions in fMRI by pressing the button with:

Pre Scan (circle one):	Left Hand	Right Hand	Both
hands			
Post Scan (circle one):	Left Hand	Right Hand	Both
hands			

2. When localizing the dot in vOICe and with the images I:

Pre-training Scan (check one):

 - Fixed my gaze on the center cross in all trials
 - Tried to fix my gaze cross but may have wandered occasionally
 - Did not try to fixate my gaze on the center cross

Post-training Scan (check one):

 - Fixed my gaze on the center cross in all trials
 - Tried to fix my gaze cross but may have wandered occasionally
 - Did not try to fixate my gaze on the center cross

3. When localizing the dot in vOICe and with the images I:

Pre-training Scan (check one):

 - Imagined pointing to the dot **after** the sound/image finished or disappeared
 - Imagined pointing to the dot **before** the sound/image finished or disappeared
 - Did not imagine pointing to the dot

Post-training Scan (check one):

 - Imagined pointing to the dot **after** the sound/image finished or disappeared
 - Imagined pointing to the dot **before** the sound/image finished or disappeared
 - Did not imagine pointing to the dot

4. When listening for a pause in the noise (just following the auditory localization)
- Pre-training Scan (check all that apply):
- I recognized that the sound was the vOICe device
 - I did not recognize that the sound was the vOICe device
 - I did not know what the vOICe device was
- Post-training Scan (check all that apply):
- I recognized that the sound was the vOICe device
 - I did not recognize that the sound was the vOICe device
 - I did not know what the vOICe device was
5. When counting backwards in sets of 7:
- Pre-training Scan (check one):
- The sound played distracted my counting significantly
 - The sound played distracted my counting somewhat
 - The sound played did not distract my counting at all
- Post-training Scan (check one):
- The sound played distracted my counting significantly
 - The sound played distracted my counting somewhat
 - The sound played did not distract my counting at all
6. When counting backwards in sets of 7:
- Pre-training Scan (check one):
- I started to imagine images of numbers
 - I counted in my head without imagining the shape or image of a number
- Post-training Scan (check one):
- I started to imagine images of numbers
 - I counted in my head without imagining the shape or image of a number
7. When listening to the natural sounds with a pause in fMRI:
- Pre-training Scan (check one):
- I started to imagine a visual scene (such as a beach)
 - I just listened to the sound for the pause with no “visual” imaginings
- Post-training Scan (check one):
- I started to imagine a visual scene (such as a beach)
 - I just listened to the sound for the pause with no “visual” imaginings

APPENDIX D

COMPLETE FMRI DATA

Chapter 4 and Chapter 5 fMRI data that were truncated to the most significant 15 peaks of neural activation are presented in full in Appendix D. The tables in Appendix D include data from Tables 4.3 (Table A in Appendix D), Table 4.4B (Table B in Appendix D), and Table 5.4A (Table C in Appendix D).

Sighted Participants (N = 10)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>Star Trek Sound [Post – Pre]</i>						
Insula	13	R	39	-46	19	0.000
Middle Temporal Gyrus	39	R	45	-55	7	0.001
<i>- small volume-corrected peak</i>						0.033*
Thalamus		R	6	-28	10	0.000
Caudate		R	21	-40	10	0.000
Thalamus		L	-6	-34	10	0.000
Middle Frontal Gyrus	6	R	33	-1	64	0.000
Caudate		R	3	5	4	0.000
Caudate		R	3	17	7	0.003
Precuneus	7	R	21	-49	46	0.000
Inferior Parietal Lobule	40	R	33	-43	46	0.001
Inferior Parietal Lobule	40	R	39	-55	46	0.004
Precentral Gyrus	6	L	-24	-16	70	0.001
Precentral Gyrus	6	L	-33	-7	67	0.005
Medial Frontal Gyrus	8	L	-12	38	34	0.001
Postcentral Gyrus	5	L	-24	-43	58	0.001
Paracentral Lobule	6	R	3	-34	70	0.002
Paracentral Lobule	4	R	9	-40	70	0.006
Lentiform Nucleus		L	-18	14	7	0.002
Caudate		L	-12	26	7	0.002
Precentral Gyrus	6	R	30	-19	70	0.003
Precentral Gyrus	4	R	42	-25	67	0.005
Superior Frontal Gyrus	8	L	-39	17	46	0.003
Middle Frontal Gyrus	8	L	-27	20	43	0.004

Sighted Participants ($N = 10$) Continued							
Region	BA	Side	x	y	z	p_{uncorr}	
Superior Frontal Gyrus		6	L	-24	14	49	0.008

Table A: The Full Version of fMRI data: post – pre training familiar sounds sighted participants (Table 4.3). Complete imaging results for sighted participants when comparing post-vOICE-training scan and the pre-vOICE-training scan ($N = 10$). All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

Late Blind Participants (N = 1) (RD)						
Region	BA	Side	x	y	z	<i>p_{uncorr}</i>
vOICE Noise Pause Detection [Post – Pre]						
Inferior Parietal Lobule	40	R	69	-25	25	0.000
<i>- small volume-corrected peak</i>						0.000*
Precentral Gyrus	4	R	60	-7	22	0.000
Supermarginal Gyrus	40	R	51	-52	25	0.000
Inferior Parietal Lobule	40	L	-60	-28	28	0.000
Supermarginal Gyrus	40	L	-48	-49	34	0.000
Supermarginal Gyrus	40	L	-42	-37	34	0.000
Middle Temporal Gyrus	39	L	-45	-67	25	0.000
<i>- small volume-corrected peak</i>						0.000*
Caudate		R	21	-1	22	0.000
Caudate		R	18	8	22	0.000
Cingulate Gyrus	24	R	24	-10	34	0.000
Superior Frontal Gyrus	8	R	18	38	52	0.000
Middle Frontal Gyrus	8	R	24	38	40	0.003
Lingual Gyrus	19	R	33	-61	1	0.000
<i>- small volume-corrected peak</i>						0.009*
Caudate		L	-15	8	19	0.000
Caudate		L	-18	-16	22	0.002
Cingulate Gyrus	24	L	-18	-19	34	0.002
Superior Parietal Lobule	7	R	36	-64	61	0.000
Postcentral Gyrus	5	R	42	-46	67	0.000
Postcentral Gyrus	2	R	42	-37	67	0.000
Superior Frontal Gyrus	6	L	-18	-13	67	0.000
Medial Frontal Gyrus	6	L	-9	-10	61	0.002

Late Blind Participants (N = 1) (RD) Continued						
Region	BA	Side	x	y	z-	<i>p</i>_{uncorr}
Superior Frontal Gyrus	6	L	-15	17	64	0.000
Precuneus	7	L	-12	-79	55	0.000
Postcentral Gyrus	5	L	-27	-40	67	0.001
Postcentral Gyrus	3	L	-30	-28	67	0.002
Lingual Gyrus	18	L	-30	-70	-8	0.001
Middle Occipital Gyrus	37	L	-36	-67	-2	0.001
Fusiform Gyrus	37	L	-36	-49	-14	0.001
Precuneus	7	L	-3	-46	52	0.002
Cingulate Gyrus	31	L	-6	-37	37	0.002
Cingulate Gyrus	31	L	0	-43	34	0.002
Cingulate Gyrus	31	R	3	-25	37	0.003
Precentral Gyrus	4	L	-54	-13	40	0.002
Superior Frontal Gyrus	9	L	-18	59	31	0.003
vOICE Distract Counting [Post – Pre]						
Middle Temporal Gyrus		R	51	-34	1	0.000
Superior Temporal Gyrus		R	63	-16	-2	0.000
Cuneus	17	R	12	-82	10	0.000
- small volume-corrected peak						0.000*
Posterior Lobe, Cerebellum		R	30	-64	-8	0.000
Posterior Lobe, Cerebellum		R	21	-76	-14	0.000
Insula	13	R	48	-22	25	0.000
Inferior Parietal Lobule	40	R	66	-37	28	0.000

Late Blind Participants (N = 1) (RD) Continued						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>- small volume-corrected peak</i>						0.000*
Inferior Parietal Lobule	40	R	39	-52	43	0.000
Middle Frontal Gyrus	8	L	-33	35	43	0.000
Middle Frontal Gyrus	8	L	-30	26	40	0.000
Middle Frontal Gyrus	9	L	-39	38	34	0.000
Inferior Parietal Lobule	40	L	-54	-28	25	0.000
Insula	13	L	-45	-19	19	0.000
Cingulate Gyrus	32	L	0	17	40	0.000
Medial Frontal Gyrus	6	L	-9	-4	58	0.000
Medial Frontal Gyrus	6	L	0	2	49	0.000
Superior Temporal Gyrus	22	L	-63	-7	4	0.000
<i>- small volume-corrected peak</i>						0.006*
Precuneus	7	L	-6	-61	43	0.000
Precuneus	7	L	-3	-79	43	0.000
Middle Frontal Gyrus	8	R	30	38	46	0.000
Superior Frontal Gyrus	9	R	39	44	34	0.002
Middle Frontal Gyrus	10	R	30	38	22	0.003
Middle Frontal Gyrus	46	R	39	26	22	0.000
Precentral Gyrus	6	R	60	-4	37	0.000
Precentral Gyrus	6	L	-51	-1	19	0.000
Middle Temporal Gyrus	39	L	-48	-58	25	0.000
Supramarginal Gyrus	40	L	-63	-49	25	0.001
Inferior Parietal Lobule	40	L	-45	-58	37	0.001
Superior Temporal Gyrus	22	L	-51	-49	7	0.000

Late Blind Participants (N = 1) (RD) Continued						
Region	BA	Side	x	y	z	<i>p_{uncorr}</i>
Claustrium		L	-27	-7	19	0.001
Caudate		L	-15	-22	19	0.001
Caudate		L	-15	-7	22	0.002
Superior Occipital Gyrus	19	R	33	-85	31	0.001
Precuneus	19	R	27	-82	43	0.002
Inferior Temporal Gyrus	20	L	-51	-55	-14	0.001
Medial Frontal Gyrus	8	L	0	53	46	0.001
Medial Temporal Gyrus	22	L	-57	-34	4	0.001
Cingulate Gyrus	31	L	0	-43	40	0.001
Precentral Gyrus	6	L	-48	-4	52	0.001
Anterior Cingulate	32	L	-18	32	19	0.001
Culmen		R	3	-49	-14	0.002
Culmen		L	-9	-43	-17	0.002
Superior Frontal Gyrus	8	R	15	44	52	0.002
Middle Frontal Gyrus	6	R	36	-4	46	0.002
Anterior Cingulate	32	L	-6	35	25	0.005
Medial Frontal Gyrus	9	L	-3	44	19	0.006
Precuneus	7	R	15	-61	37	0.005
Cuneus	19	R	15	-79	31	0.005
Precuneus	7	R	21	-67	31	0.006
Beach Pause Detection [Post – Pre]						
Precuneus	19	L	-24	-85	43	0.000
- small volume-corrected peak						0.000*
Supramarginal Gyrus	40	L	-60	-46	37	0.000

Late Blind Participants (N = 1) (RD) Continued						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>- small volume-corrected peak</i>						0.000*
Superior Occipital Gyrus	19	L	-36	-82	34	0.000
Middle Temporal Gyrus	39	R	45	-61	28	0.000
<i>- small volume-corrected peak</i>						0.001*
Inferior Parietal Lobule	40	R	69	-25	25	0.000
Precuneus	19	R	33	-79	34	0.000
Middle Frontal Gyrus	8	L	-45	17	49	0.000
Superior Frontal Gyrus	8	L	-27	44	40	0.000
Superior Frontal Gyrus	9	L	-18	59	34	0.005
Superior Frontal Gyrus	9	L	-27	56	34	0.008
Superior Frontal Gyrus	10	L	-42	50	25	0.000
Lingual Gyrus	19	L	-33	-67	-2	0.004
Star Trek Pause Detection [Post – Pre]						
Cuneus	17	R	9	-82	10	0.000
<i>- small volume-corrected peak</i>						0.000*
Lingual Gyrus	18	L	-15	-79	-5	0.000
<i>- small volume-corrected peak</i>						0.003*
Lingual Gyrus	18	R	18	-70	4	0.000
Superior Temporal Gyrus	39	R	48	-55	25	0.000
<i>- small volume-corrected peak</i>						0.000*
Inferior Parietal Lobule	40	R	69	-31	28	0.000
<i>- small volume-corrected peak</i>						0.000*
Postcentral Gyrus	2	R	45	-25	31	0.000
Middle Temporal Gyrus	39	L	-42	-61	25	0.000

Late Blind Participants ($N = 1$) (RD) Continued						
Region	BA	Side	x	y	z	p_{uncorr}
Inferior Parietal Lobule	40	L	-57	-28	25	0.000
Inferior Parietal Lobule	40	L	-48	-34	28	0.000
Precuneus	19	R	33	-79	34	0.001
- small volume-corrected peak						0.044*
Precuneus	7	L	-21	-79	49	0.002

Table B: The Full Version of fMRI data: post – pre training late blind participant (Table 4.4 B). Complete imaging results for a late blind participant ($N = 1$) when comparing post-vOICe-training scan and the pre-vOICe-training scan. All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold (p_{uncorr} refers to the peak level p_{uncorr}). The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

Late Blind Participants (N = 1) (RD)						
Region	BA	Side	x	y	z	<i>p</i>_{uncorr}
<i>vOICe Dot Post [Right – Left]</i>						
No Activation						
<i>vOICe Dot Post [Left – Right]</i>						
Fusiform Gyrus	37	R	42	-55	-8	0.000
Clastrum			36	-22	-2	0.000
Fusiform Gyrus	19	R	42	-73	-11	0.000
Temporal Lobe	37	L	-42	-46	-8	0.000
Culmen		L	-18	-58	-8	0.000
Culmen		L	-21	-49	-11	0.000
Cuneus	18	R	15	-67	16	0.000
Posterior Cingulate	30	R	15	-52	13	0.000
Cuneus	18	R	12	-76	25	0.000
Middle Temporal Gyrus	39	R	51	-76	25	0.000
Middle Temporal Gyrus	39	R	57	-67	25	0.000
Middle Temporal Gyrus	39	R	60	-64	13	0.003
Middle Occipital Gyrus	18	L	-24	-82	-8	0.000
Thalamus		L	-3	-7	10	0.000
Lentiform Nucleus		L	-18	2	10	0.002
Inferior Frontal Gyrus	45	R	57	14	22	0.001
Middle Occipital Gyrus	19	L	-36	-70	13	0.001
Insula	13	R	39	-4	19	0.002
Middle Temporal Gyrus	21	L	-51	-31	-5	0.002
Clastrum		R	36	2	7	0.002
Inferior Frontal Gyrus	45	L	-57	17	19	0.003

Late Blind Participants ($N = 1$) (RD) Continued						
Region	BA	Side	x	y	z	p_{uncorr}
Lentiform Nucleus		R	18	5	10	0.005
Lentiform Nucleus		R	21	2	1	0.005

Table C. The Full Version of fMRI data: vOICE dot [Right – Left location] post-scan late blind participant (Table 5.4 A). Complete imaging results for a late blind participant when comparing the post-training left dot and the post-training right dot in vOICE ($N = 1$). All regions were limited to $p < 0.009$ uncorrected and 10 voxel cluster threshold. The small volume correction was for a sphere of 10 millimeter radius around the cluster center, and the pvalue shown (indicated by asterisk, *i.e.*, *) is for the peak level FWE-corrected. Brodmann Area localization was performed on the talarach client for nearest grey matter. Any clusters without nearest grey matter within +/- 5 mm are not included.

REFERENCES

- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839-843.
- Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., . . . Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*(6), 687-689.
- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*(3), 301-308.
- Araque, N. O., Dunai, L., Rossetti, F., Listl, L., Mirmehdi, M., Mora, J. L. G., . . . Dunai, I. (2008). Sound Map Generation for a Prototype Blind Mobility System Using Multiple Sensors. *Service Robotics & Smart Homes: How a gracefully adaptive integration of both environments can be envisaged?*
- Arno, P., De Volder, A. G., Vanlierde, A., Wanet-Defalque, M. C., Streel, E., Robert, A., . . . Veraart, C. (2001). Occipital activation by pattern recognition in the early blind using auditory substitution for vision. *Neuroimage*, *13*(4), 632-645.
- Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., . . . Stephan, K. (2011). SPM8 manual. London: Functional Imaging Laboratory, Wellcome Trust Centre for Neuroimaging.
- Auvray, M., Hanne-ton, S., Lenay, C., & O'Regan, K. (2005). There is something out there: distal attribution in sensory substitution, twenty years later. *Journal of Integrative Neuroscience*, *4*(04), 505-521.

- Auvray, M., Hanneton, S., & O Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with the vOICE. *Perception, 36*(3), 416.
- Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection.
- Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., & Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: A technical note. *Journal of Rehabilitation Research and Development, 35*(4), 427-430.
- Baseler, H. A., Brewer, A. A., Sharpe, L. T., Morland, A. B., Jagle, H., & Wandell, B. A. (2002). Reorganization of human cortical maps caused by inherited photoreceptor abnormalities. *Nature Neuroscience, 5*(4), 364-370.
- Baseler, H. A., Gouws, A., Haak, K. V., Racey, C., Crossland, M. D., Tufail, A., . . . Morland, A. B. (2011). Large-scale remapping of visual cortex is absent in adult humans with macular degeneration. *Nature Neuroscience, 14*(5), 649-655.
- Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nature Reviews Neuroscience, 3*(6), 443-452.
- Beauchamp, M., Argall, B., Bodurka, J., Duyn, J., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7*(11), 1190-1192.
- Beauchamp, M., Lee, K., Argall, B., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809-823.

- Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919), 696-698.
- Bernal, B., & Perdomo, J. Brodmann's Interactive Atlas, Area 39 Retrieved March 31, 2014, from <http://www.fmriconsulting.com/brodmann/BA39.html>
- Bernal, B., & Perdomo, J. (2014, 2008). Brodmann's Interactive Atlas, Directory of Function Retrieved 5/23/2014, 2014, from <http://www.fmriconsulting.com/brodmann/functions.html>
- Bittar, R. G., Ptito, M., Faubert, J., Dumoulin, S. O., & Ptito, A. (1999). Activation of the remaining hemisphere following stimulation of the blind hemifield in hemispherectomized subjects. *Neuroimage*, *10*(3), 339-346.
- Boroojerdi, B., Bushara, K. O., Corwell, B., Immisch, I., Battaglia, F., Muellbacher, W., & Cohen, L. G. (2000). Enhanced excitability of the human visual cortex induced by short-term light deprivation. *Cerebral Cortex*, *10*(5), 529-534.
- Bouvrie, J. V., & Sinha, P. (2007). Visual object concept discovery: Observations in congenitally blind children, and a computational approach. *Neurocomputing*, *70*(13-15), 2218-2233.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244.
- Browne, R. (2003). Toward a mobility aid for the blind. *Proc. Image and Vision Computing*.

- Burton, H., Snyder, A. Z., Conturo, T. E., Akbudak, E., Ollinger, J. M., & Raichle, M. E. (2002). Adaptive changes in early and late blind: a fMRI study of Braille reading. *Journal of Neurophysiology*, *87*(1), 589-607.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649-657.
- Capelle, C., Trullemans, C., Arno, P., & Veraart, C. (2002). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions on Biomedical Engineering*, *45*(10), 1279-1293.
- Cattaneo, Z., & Vecchi, A. (2011). *Blind vision: the neuroscience of visual impairment*. Cambridge, Massachusetts: MIT Press.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200-219.
- Changizi, M. A., Hsieh, A., Nijhawan, R., Kanai, R., & Shimojo, S. (2008). Perceiving the present and a systematization of illusions. *Cognitive Science*, *32*(3), 459-503.
- Chebat, D. R., Schneider, F. C., Kupers, R., & Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport*, *22*(7), 342.
- Chekhchoukh, A., Vuillerme, N., & Glade, N. (2011). *Vision substitution and moving objects tracking in 2 and 3 dimensions via vectorial electro-stimulation of the tongue*. Paper presented at the Actes de ASSISTH 2011, 2eme Conference internationale sur l'Accessibilite et les Systemes de Suppleance aux personnes en situations de Handicaps, Paris, France.

- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25, 975.
- Cheung, S., Fang, F., He, S., & Legge, G. E. (2009). Retinotopically specific reorganization of visual cortex for tactile pattern recognition. *Current Biology*, 19(7), 596-601.
- Clemons, J., Bao, S. Y., Savarese, S., Austin, T., & Sharma, V. (2012). *MVSS: Michigan Visual Sonification System*. Paper presented at the 2012 IEEE International Conference on Emerging Signal Processing Applications (ESPA), Las Vegas, Nevada.
- Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Faiz, L., Dambrosia, J., . . . Catala, M. D. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature*, 389(6647), 180-183.
- Collignon, O., Lassonde, M., Lepore, F., Bastien, D., & Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cerebral Cortex*, 17(2), 457.
- Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental Brain Research*, 192(3), 343-358.
- Corbett, J. E., Enns, J. T., & Handy, T. C. (2009). Electrophysiological evidence for a post-perceptual influence of global visual context on perceived orientation. *Brain Research*, 1292, 82-92.

- Crutch, S. J., Lehmann, M., Gorgoraptis, N., Kaski, D., Ryan, N., Husain, M., & Warrington, E. K. (2011). Abnormal visual phenomena in posterior cortical atrophy. *Neurocase*, *17*(2), 160-177.
- Day, R. H. (1968). Perceptual constancy of auditory direction with head rotation. *Nature*, *219*, 501-502.
- Dayan, P., Abbott, L. F., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge: MIT Press.
- Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic, an fMRI study. *Cognitive Brain Research*, *18*(1), 76-88.
- Denney, D., & Adorjanti, C. (1972). Orientation specificity of visual cortical neurons after head tilt. *Experimental Brain Research*, *14*(3), 312-317.
- Dunai, L. (2010). Design, modeling and analysis of object localization through acoustical signals for cognitive electronic travel aid for blind people. *Universidad Politecnica De Valencia, School of Design Engineering, Ph.D. Thesis*.
- Eramudugolla, R., Kamke, M. R., Soto-Faraco, S., & Mattingley, J. B. (2011). Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition*, *118*(1), 62-74.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5).
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

- Farah, M., Soso, M., & Dasheiff, R. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 241.
- Fujii, T., Tanabe, H. C., Kochiyama, T., & Sadato, N. (2009). An investigation of cross-modal plasticity of effective connectivity in the blind by dynamic causal modeling of functional MRI data. *Neuroscience Research*, 65(2), 175-186.
- Gibson, J. J. (1950a). The perception of the visual world.
- Gibson, J. J. (1950b). *The perception of the visual world*. Boston: Houghton Mifflin.
- Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, 22(5), 383-388.
- Haigh, A., Brown, D. J., Meijer, P., & Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in Psychology*, 4.
- Hamilton, R., Keenan, J. P., Catala, M. D., & Pascual-Leone, A. (2000). Alexia for Braille following bilateral occipital stroke in an early blind woman. *Neuroreport*, 11(02), 237-240.
- Hannula, D. E., Simons, D. J., & Cohen, N. J. (2005). Imaging implicit perception: promises and pitfalls. *Nature Reviews Neuroscience*, 6(3), 247-255.
- Harris, C. S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review*, 72(6), 419.
- Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision*, 8(1).

- Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56(5), 872.
- Henriksson, L., Raninen, A., Nasanen, R., Hyvarinen, L., & Vanni, S. (2007). Training-induced cortical representation of a hemianopic hemifield. *Journal of Neurology, Neurosurgery and Psychiatry*, 78(1), 74-81.
- Humayun, M. S., Weiland, J. D., Fujii, G. Y., Greenberg, R., Williamson, R., Little, J., . . . Dagnelie, G. (2003). Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Research*, 43(24), 2573-2581.
- Inui, T., Otsu, Y., Tanaka, S., Okada, T., Nishizawa, S., & Konishi, J. (1998). A functional MRI analysis of comprehension processes of Japanese sentences. *Neuroreport*, 9(14), 3325-3328.
- James, T. W., & Stevenson, R. A. (2012). The Use of fMRI to Assess Multisensory Integration. In M. M. M. & W. M. T. (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton: CRC Press.
- Kaski, D. (2002). Revision: Is visual perception a requisite for visual imagery? *Perception*, 31(6), 717-732.
- Klein, I., Dubois, J., Mangin, J. F., Kherif, F., Flandin, G., Poline, J. B., . . . Le Bihan, D. (2004). Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Cognitive Brain Research*, 22(1), 26-31.
- Klinge, C., Eippert, F., Roder, B., & Buchel, C. (2010). Corticocortical connections mediate primary visual cortex responses to auditory stimulation in the blind. *The Journal of Neuroscience*, 30(38), 12798-12805.

- Kohler, S., Kapur, S. j., Moscovitch, M., Winocur, G., & Houle, S. (1995). Dissociation of pathways for object and spatial vision: a PET study in humans. *Neuroreport*, 6(14), 1865-1868.
- Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of Neuroscience*, 30(29), 9801-9820.
- Kosslyn, S., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Ganis, G., . . . Alpert, N. M. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, 284(5411), 167-170.
- Kosslyn, S., & Thompson, W. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129(5), 723-746.
- Kosslyn, S., Thompson, W., Klm, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378, 496-498.
- Kourtzi, Z., Erb, M., Grodd, W., & Bulthoff, H. H. (2003). Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cerebral Cortex*, 13(9), 911-920.
- Kujala, T., Huotilainen, M., Sinkkonen, J., Ahonen, A. I., Alho, K., Ilmoniemi, R. J., . . . Salonen, O. (1995). Visual cortex activation in blind humans during sound discrimination. *Neuroscience Letters*, 183(1), 143-146.
- Kupers, R., Chebat, D. R., Madsen, K. H., Paulson, O. B., & Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *Proceedings of the National Academy of Sciences*, 107(28), 12716-12721.

- Kupers, R., Fomal, A., de Noordhout, A. M., Gjedde, A., Schoenen, J., & Ptito, M. (2006). Transcranial magnetic stimulation of the visual cortex induces somatotopically organized qualia in blind subjects. *Proceedings of the National Academy of Sciences*, *103*(35), 13256-13260.
- Lambert, S., Sampaio, E., Mauss, Y., & Scheiber, C. (2004). Blindness and brain plasticity: contribution of mental imagery?: An fMRI study. *Cognitive Brain Research*, *20*(1), 1-11.
- Linden, D., Kallenbach, U., Heinecke, A., Singer, W., & Goebel, R. (1999). The myth of upright vision. A psychophysical and functional imaging study of adaptation to inverting spectacles. *Perception*, *28*, 469-481.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman and Company.
- Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, *39*(2), 112-121.
- Merabet, L. B., Battelli, L., Obretenova, S., Maguire, S., Meijer, P., & Pascual-Leone, A. (2009). Functional recruitment of visual cortex for sound encoded object identification in the blind. *NeuroReport*, *20*(2), 132-138. doi: 10.1097/WNR.0b013e32832104dc
- Merabet, L. B., Rizzo, J. F., Amedi, A., Somers, D. C., & Pascual-Leone, A. (2005). What blindness can tell us about seeing again: merging neuroplasticity and neuroprostheses. *Nature Reviews Neuroscience*, *6*(1), 71-77.

- Milne, J. L., Goodale, M. A., & Thaler, L. (2013). Is there a 'retinotopic' representation of echo locations in the calcarine cortex of the blind brain? *Journal of Vision, 13*(9), 1334-1334.
- Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *The Journal of Neuroscience, 27*(15), 4120-4131.
- Miyauchi, S., Egusa, H., Amagase, M., Sekiyama, K., Imaruoka, T., & Tashiro, T. (2004). Adaptation to left-right reversed vision rapidly activates ipsilateral visual cortex in humans. *Journal of Physiology, 98*(1), 207-219.
- Morgan, G. A., Goodson, F. E., & Jones, T. (1975). Age differences in the associations between felt temperatures and color choices. *The American Journal of Psychology, 125*-130.
- Mountcastle, V. B. (1978). Brain mechanisms for directed attention. *Journal of the Royal Society of Medicine, 71*(1), 14.
- Neri, P., & Levi, D. M. (2007). Temporal dynamics of figure-ground segregation in human vision. *Journal of Neurophysiology, 97*(1), 951-957.
- Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Research, 405*(2), 268-283.
- Neville, H. J., Schmidt, A., & Kutas, M. (1983). Altered visual-evoked potentials in congenitally deaf adults. *Brain Research, 266*(1), 127-132.

- Niimi, R., Saneyoshi, A., Abe, R., Kaminaga, T., & Yokosawa, K. (2011). Parietal and frontal object areas underlie perception of object orientation in depth. *Neuroscience Letters*, *496*(1), 35-39.
- O'Regan, J. K., & Noe, A. (2001). What it is like to see: A sensorimotor theory of perceptual experience. *Synthese*, *129*(1), 79-103.
- Ortiz, T., Poch, J., Santos, J. M., Requena, C., Martinez, A. M., Ortiz-Teran, L., . . . Calvo, A. (2011). Recruitment of occipital cortex during sensory substitution training linked to subjective experience of seeing in people with blindness. *PLoS One*, *6*(8), e23264.
- Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, *45*(1), 75-92.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*: MIT Press Cambridge, MA.
- Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, *220*(3-4), 319-333.
- Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, *134*, 427-445.
- Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology*, *15*(2), 188-196.
- Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458-11463.

- Plaza, P., Cuevas, I., Collignon, O., Grandin, C., De Volder, A., & Renier, L. (2009, June 29, 2009). *Perceiving schematic faces and man-made objects by a visual-to-auditory sensory substitution activates the fusiform gyrus*. Paper presented at the 10th International Multisensory Research Forum (IMRF), New York City.
- Plaza, P., Cuevas, I., Grandin, C., De Volder, A. G., & Renier, L. (2012). Looking into Task-Specific Activation Using a Prosthesis Substituting Vision with Audition. *ISRN Rehabilitation, 2012*.
- Poirier, C., De Volder, A., Tranduy, D., & Scheiber, C. (2007). Pattern recognition using a device substituting audition for vision in blindfolded sighted subjects. *Neuropsychologia, 45*(5), 1108-1121.
- Poirier, C., De Volder, A. G., & Scheiber, C. (2007). What neuroimaging tells us about sensory substitution. *Neuroscience & Biobehavioral Reviews, 31*(7), 1064-1070.
- Poirier, C., De Volder, A. G., Tranduy, D., & Scheiber, C. (2006). Neural changes in the ventral and dorsal visual streams during pattern recognition learning. *Neurobiology of Learning and Memory, 85*(1), 36-43.
- Poirier, C., Richard, M., Duy, D. T., & Veraart, C. (2006). Assessment of sensory substitution prosthesis potentialities in minimalist conditions of learning. *Applied Cognitive Psychology, 20*(4), 447-460.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3-25.
- Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology, 13*(3), 278.

- Proulx, M. J., Stoerig, P., Ludowig, E., & Knoll, I. (2008). Seeing where through the ears: effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution. *PloS One*, 3(3).
- Ptito, M., Moesgaard, S. M., Gjedde, A., & Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain*, 128(3), 606-614.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102-1107.
- Ramachandran, V. S., & Hubbard, E. M. (2003). Hearing colors, tasting shapes. *Scientific American*, 288(5), 52-59.
- Rauschecker, J. P., & Korte, M. (1993). Auditory compensation for early blindness in cat cerebral cortex. *The Journal of Neuroscience*, 13(10), 4538-4548.
- Renier, L., Bruyer, R., & De Volder, A. G. (2006). Vertical-horizontal illusion present for sighted but not early blind humans using auditory substitution of vision. *Perception & Psychophysics*, 68(4), 535.
- Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., . . . De Volder, A. G. (2005). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *NeuroImage*, 26(2), 573-580.
- Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., . . . Devolder, A. (2005). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *NeuroImage*, 26(2), 573-580. doi: 10.1016/j.neuroimage.2005.01.047

- Renier, L., & De Volder, A. G. (2010). Vision substitution and depth perception: Early blind subjects experience visual perspective through their ears. *Disability & Rehabilitation: Assistive Technology*, 5(3), 175-183.
- Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., & De Volder, A. G. (2005). The Ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception*, 34(7), 857-867.
- Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. P., & Mariotti, S. P. (2004). Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization*, 82(11), 844-852.
- Reynolds, Z., & Glenney, B. (2012). When Sensory Substitution Devices Strike Back: An Interactive Training Paradigm. *Philosophy Study*, 2(6), 451-457.
- Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, J. V., . . . Pietrini, P. (2009). Do we really need vision? How blind people „Äusee,Äù the actions of others. *The Journal of Neuroscience*, 29(31), 9719-9724.
- Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. *Cognitive Psychology*, 24(4), 502-534.
- Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography*, 21(3-4), 185-192.
- Rosler, F., Roder, B., Heil, M., & Hennighausen, E. (1993). Topographic differences of slow event-related brain potentials in blind and sighted adult human subjects during haptic mental rotation. *Cognitive Brain Research*, 1(3), 145-159.

- Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(6574), 526-528.
- Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J., & Driver, J. (2005). Attentional load and sensory competition in human vision: modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. *Cerebral Cortex*, *15*(6), 770-786.
- Sereno, M., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., . . . Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889-893.
- Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545), 1350-1354.
- 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.
- Shimojo, S. (2008). Self and world: large scale installations at science museums. *Spatial Vision*, *21*(3-5), 337-346.
- Shimojo, S., & Nakajima, Y. (1981). Adaptation to the reversal of binocular depth cues: effects of wearing left-right reversing spectacles on stereoscopic depth perception. *Perception*, *10*(4), 391.
- Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, *11*(4), 505-509.

- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317-1322.
- Shimony, J. S., Burton, H., Epstein, A. A., McLaren, D. G., Sun, S. W., & Snyder, A. Z. (2006). Diffusion tensor imaging reveals white matter reorganization in early blind humans. *Cerebral Cortex*, *16*(11), 1653-1661.
- Shu, N., Liu, Y., Li, J., Li, Y., Yu, C., & Jiang, T. (2009). Altered anatomical network in early blindness revealed by diffusion tensor tractography. *PLoS One*, *4*(9), e7228.
- Slotnick, S., Thompson, W., & Kosslyn, S. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, *15*(10), 1570-1583.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, *22*(1), 245-260.
- Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences of the United States of America*, *54*(2), 407.
- Stiles, N. R. B., McIntosh, B. P., Nasiatka, P. J., Hauer, M. C., Weiland, J. D., Humayun, M. S., & Tanguay Jr, A. R. (2011). An intraocular camera for retinal prostheses: Restoring sight to the blind. In A. Serpenguzel & A. W. Poon (Eds.), *Optical Processes in Microparticles and Nanostructures: A Festschrift Dedicated to Richard Kounai Chang on His Retirement from Yale University* (pp. 385): World Scientific Publishing Co.

- Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron*, *76*(3), 640-652.
- Sugita, Y. (1996). Global plasticity in adult visual cortex following reversal of visual input. *Nature*, *380*(6574), 523-526.
- Tootell, R., Silverman, M., Switkes, E., & De Valois, R. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, *218*(4575), 902-904.
- Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., . . . Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, *15*(4), 3215-3230.
- Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156-177.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242-248.
- Uhl, F., Franzen, P., Lindinger, G., Lang, W., & Deecke, L. (1991). On the functionality of the visually deprived occipital cortex in early blind persons. *Neuroscience Letters*, *124*(2), 256-259.
- Uhl, F., Kretschmer, T., Lindinger, G., Goldenberg, G., Lang, W., Oder, W., & Deecke, L. (1994). Tactile mental imagery in sighted persons and in patients suffering from peripheral blindness early in life. *Electroencephalography and Clinical Neurophysiology*, *91*(4), 249-255.

Visual Impairment and Blindness. (2012). *Fact Sheet Number 282*. Retrieved from World Health Organization website:

<http://www.who.int/mediacentre/factsheets/fs282/en/index.html>

von Helmholtz, H. (1925). *The Perceptions of Vision* (Vol. III): Optical Society of America.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, 30(3), 829-841.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience*, 5(5), 491-499.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21-25.

Walker, P., Francis, B. J., & Walker, L. (2010). The brightness-weight illusion: Darker objects look heavier but feel lighter. *Experimental Psychology*, 57(6), 462.

Ward, J., & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, 19(1), 492-500.

Ward, J., & Wright, T. (2014). Sensory substitution as an artificially acquired synaesthesia. *Neuroscience and Biobehavioral Reviews*, 41, 26-35.

- Winter, J. O., Cogan, S. F., & Rizzo, J. F. (2007). Retinal prostheses: current challenges and future outlook. *Journal of Biomaterials Science, Polymer Edition*, 18(8), 1031-1055.
- Zahorik, P., & Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience*, 4(1), 78-83.
- Zangenehpour, S., & Zatorre, R. J. (2010). Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, 48(2), 591-600.