# Sound localization training with audio-visual integration in sub-optimal listening environments

By

Michael Kiewe

A dissertation submitted in partial fulfillment of

the requirement for the degree of

Doctor of Philosophy

(Physics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2014

Defended on 28 July 2014

Dissertation approved by the following members of the Final Oral Committee:

Ruth Litovsky, Professor, Communicative Disorders

Susan Coppersmith, Professor, Physics

Michael Winokur, Professor, Physics

Deniz Yavuz, Professor, Physics

James Titus, Post-Doctoral Fellow, Physics

# ABSTRACT

Cochlear implant (CI) users are outperformed by normal hearing (NH) listeners when localizing sound sources in space [1–3]. This may result from inaccurate representation of auditory space, due to lack of experience in integrating auditory cues that are unreliable with visual cues that are highly reliable and well-mapped in the spatial dimension.

A training method was designed in order to assess whether sound localization performance could be improved, by reinforcing salient localization cues over degraded localization cues. This training method is hypothesized to allow the listener to weight the former more heavily than the latter. This approach was developed and tested by training 23 NH adults, in two sub-optimal listening environments with conflicting auditory cues.

The audio-visual integration training (AVIT) method relies on a visual stimulus that coincides with the sound's location, serving to reinforce a more fine-tuned auditory spatial map by allowing the listener to optimally weight the available auditory cues. In the first experiment, the sub-optimal listening environment was echoic; listeners were trained to weight the auditory cues associated with the first-arriving soundwave more heavily than the later-arriving soundwave. In the second experiment, vocoded stimuli were used to induce a sub-optimal listening situation, in which CI processing is modeled and used to simulate realistic conditions experienced by CI users. NH listeners were trained to weight relatively intact interaural level difference cues more heavily than other cues that were degraded by the vocoding process. AVIT was shown to produce improvement in sound localization ability in the first experiment, suggesting that degradation of spatial cues by echoes can be overcome with audio-visual integration, whereas, degradation of spatial cues with CI processing is more challenging to overcome.

# ACKNOWLEDGEMENTS

I would like to sincerely thank all of the people who took time to make this work possible. First, thank you to my advisor, Ruth Litovsky, who guided me through my research, gave me the opportunity to improve my teaching and served as a source of inspiration for me. Your support have been instrumental in shaping my future and I am truly appreciative of it.

Thank you to all the members of the Binaural Hearing and Speech Lab who were the best co-workers I could possibly hope for. I especially want to thank my lab manager Shelly Godar for always being cheerful and helping me deal the many logistics hurdles involved in building an experiment, recruiting subjects for it and testing them. You and Ruth constantly impress me with the way you manage to juggle so many things at once.

To my office mate Heath Jones, whom I've had many meaningful conversations about research and life, thank you for being a role model for me (even if you don't know it). You are truly a person with a solid sense of self and I hope to one day be as successful as you at making other people feel welcomed and comfortable.

Thank you to Alan Kan, the senior BHSL member, who guided me through the programming, engineering, and theory aspects of my project. You made me comfortable attempting things I thought I could not do and imparted upon me the importance of being independent.

Thank you to my co-worker and friend, Ann Todd, who made me laugh the most at work and whose hard work ethic pushed me to do more.

I would like to thank all those in the Physics department who provided support to me during my time at UW-Madison. Thank you to Jim Reardon who helped shape my view on teaching. I've had many interesting and fun conversations with you and still you are an enigma to me. You also helped me make some important decisions regarding my future and I am forever thankful for that.

Thank you to Jane Schimmel and Renee Lefkow, the Graduate Student Coordinators, who were always friendly and willing to help. You were my go to ladies.

Thank you my Physics department advisor, Sue Coppersmith, who stepped up when I needed her. You came to my defense in my time of need and I appreciate that greatly.

Thank you to Nick Balster, Susan Nossal and Larry Watson who also help shape my view as a teacher. You gave me so many tools and knowledge that will undoubtedly help me be a successful teacher.

Thank you to the undergraduate students taking physics courses who were the participants in my "teaching experiments". You allowed me to become a better teacher, sometimes at your expense.

I would also like to thank my family and friends for all the support they have given me. First to my Israeli friends: Yevgeniy Katz, thank you for keeping in touch all these years. I will never forget our time together. "flash and go" my friend "flash and go". To Dima Mikitchuk, you've taught me a very valuable skill – the ability to enjoy misery. Though times of misery are rare these days for me, I cannot express how grateful I am for what you've given me as it got me through some difficult and lonely times. I admire your unshakable inner peace and I will never forget you.

To the many friends I’ve met while in Madison. You’ve made my time here extremely enjoyable. I can honestly say that each year was better than the one before it. You have allowed me to learn so much about myself. I have grown socially and emotionally because of you. You’ve made me a better person and I thank you for that. I will always cherish the memories of our crazy adventures.

# Contents

# LIST OF FIGURES

# LIST OF TABLES

# I. INTRODUCTION

The aim of this work was to develop an optimal training method that could improve the sound localization ability of people who use bilateral cochlear implants. The main hypothesis was that a visual stimulus serving as a salient localization cue could help induce an optimal weighting of the different degraded localization cues available to a cochlear implant (CI) user. The developed method was tested with normal hearing (NH) listeners under conditions that simulate the difficulties faced by CI users.

For a NH person, an incident sound wave is amplified and filtered at the outer ear. The air pressure vibrations are then carried over to mechanical vibrations and amplified in the middle ear. The mechanical vibrations carry over into the inner ear in the form of fluid displacement inside the cochlea, which triggers mechanical vibrations in the specialized part of the cochlea called the organ of corti. There the mechanical vibrations displace piezoelectric sensors called stereocilia. The resulting electric signal then propagates to the auditory nerve and up the auditory system, eventually arriving at the cortex where perception occurs.

For a CI user, an incident sound wave is picked up by a microphone worn by the individual. The signal is then sent to a processor where it is compressed and bandpass filtered into different frequency channels. In each channel the temporal envelope is extracted and used to modulate a carrier signal made of biphasic pulses. The resulting signal is then transmitted via radio-frequency link to a subcutaneous receiver and is sent in the form of electric current to electrodes situated along the cochlea. The current stimulates the auditory nerve directly and from the auditory nerve the signal continues through the auditory pathway.

The outline of this work is as follows. The remainder of Chapter I presents the background necessary to understand the difficulties faced by CI users when localizing sounds, the relevant factors to constructing an optimal training method that is also tailored to these difficulties, and the ways to test this method with NH subjects. This chapter motivates the choices made when designing the experiments to test the hypothesis. Chapters II and III describe the two experiments used to test the efficacy of the aforementioned training method, as well as their results. Chapter IV interprets the results with respect to the main hypothesis, compares the results with similar studies from the literature and debates the prospect of the training method to be used as a rehabilitation tool for people with CIs.

## A. Sound Localization

A NH person uses 3 main auditory cues to locate sound in space, namely interaural level differences (ILDs), interaural timing differences (ITDs) and spectral cues. ITDs occur when a sound has to travel a longer distance to one ear compared to the other. ILDs are affected by the sound's interaction with the listener's head and thus are frequency dependent. ITDs can be present when comparing the interaural fine structures and when comparing the interaural envelopes of high-frequency carriers. These binaural cues enable localization in the horizontal plane (azimuth). Spectral cues, resulting from the sound wave's interaction with the head and pinnae, are associated with localization in elevation and differentiating sounds sources originating from front and back. Spectral cues are essentially level cues however they can be both interaural and monaural [4,5].

A bilateral CI (BiCI) user is subjected to a degraded version of the aforementioned localization cues. The absence of temporal fine structure in the CI processors, and the lack of synchronization between the two CIs, limit and degrade ITD information present in the signal [6,7]. Degradation of ILDs occurs due to the CI's compression function which attenuates high amplitude sounds and amplifies low amplitude sounds [8]. Finally, spectral cues associated with the pinnae are typically not available (with some exceptions), because clinical practice involves placing the CI microphone behind the ear. Moreover, spectral cues associated with the head are degraded due to the poor frequency resolution associated with both the sound processing procedure and overlapping regions of excitation associated with the electrodes [9,10].

### A.1. Localization error types

Hartmann et al. (1998) described a choice of statistics used to quantify localization errors by defining the root mean square (RMS) error that encompasses two different types of errors: The variability error that described the listener's precision and the bias error that describes the listener's accuracy. The exact definition for these errors are presented in appendix A for the case of equal presentation number for each target location as in this study and most other localization studies.

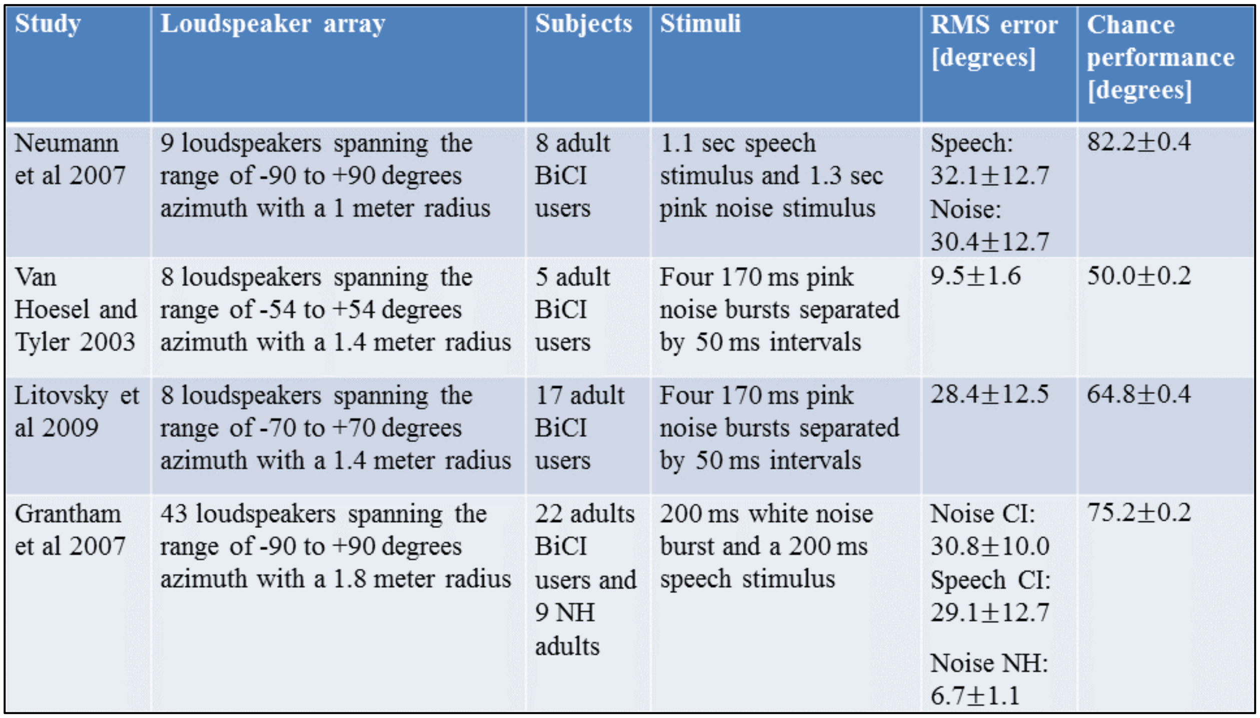
Error type can point to the decision strategy of the listener. When a listener encounters a sound and is unsure about its origin, they are usually unsure within a certain area in space (as opposed to assigning equal probability to each point in space). A listener who exhibits poor precision and good accuracy might be responding somewhat randomly within that area, whereas a listener who exhibits good precision and poor accuracy might be choosing one location within that area and repeatedly responding as if the sound is coming from that location. The former listener's strategy may be representing bottom-up effects – essentially how salient is the spatial information in the sound and how efficient are the auditory periphery in transmitting this information to higher areas in the brain where perception occurs. The latter listener's strategy may be representing top-down effects – essentially how higher areas in the brain (for which the neurological model is much less clear than that for the periphery) affect the transmitted signal before perception occurs. Therefore, delineating the type of error in localization studies could potentially help improve the neurological model for different areas in the brain.

### A.2. Typical localization error values for NH listeners and BiCI users

The specific structure of loudspeaker arrays used for localization studies influences the localization error values reported in various studies. An array that spans a large angle range enables the listener to make larger maximum errors. An array that has a large angle between neighboring loudspeakers enables the listener to make larger minimum errors. Table 1 shows RMS errors for BiCI users as well as NH listeners from different published studies.

The values differ greatly from one study to the other due to two main factors. The influence of the loudspeaker array in each study can be seen from the different *chance performance* values in the table. These are the average RMS errors for a listener that performs the localization task by responding at random. Grantham et al. (2007) showed that bilaterally deaf subjects listening with one CI localized sounds at chance performance, thereby demonstrating the importance of using BiCIs. The second factor has to do with the large intersubject variability, a well-known hallmark of studies with CI users [12,13]. CI users differ from one another in the cause of deafness, duration of deafness prior to implantation, age at implantation, delay between implantation of the first CI and the second CI, number of months post CI activation at time of study, number of electrodes in CI, processing strategy in CI, amount of neural degeneration due to auditory deprivation, age and other factors that are not well understood. The difficulty of assessing the typical localization ability of a BiCI user is exacerbated further by the small number of subjects in some studies [14]. Nevertheless, there are two main conclusions one could take from table 1. First, BiCI users perform better than chance performance. Second, BiCI users perform much worse than NH listeners.

Table 1. Representative studies of localization ability by NH listeners and CI users quantified by RMS error. The chance performance category shows the average RMS error for random responses.



| Study | Loudspeaker array | Subjects | Stimuli | RMS error [degrees] | Chance performance [degrees] |
|---|---|---|---|---|---|
| Neumann et al 2007 | 9 loudspeakers spanning the range of -90 to +90 degrees azimuth with a 1 meter radius | 8 adult BiCI users | 1.1 sec speech stimulus and 1.3 sec pink noise stimulus | Speech: 32.1±12.7<br>Noise: 30.4±12.7 | 82.2±0.4 |
| Van Hoesel and Tyler 2003 | 8 loudspeakers spanning the range of -54 to +54 degrees azimuth with a 1.4 meter radius | 5 adult BiCI users | Four 170 ms pink noise bursts separated by 50 ms intervals | 9.5±1.6 | 50.0±0.2 |
| Litovsky et al 2009 | 8 loudspeakers spanning the range of -70 to +70 degrees azimuth with a 1.4 meter radius | 17 adult BiCI users | Four 170 ms pink noise bursts separated by 50 ms intervals | 28.4±12.5 | 64.8±0.4 |
| Grantham et al 2007 | 43 loudspeakers spanning the range of -90 to +90 degrees azimuth with a 1.8 meter radius | 22 adults BiCI users and 9 NH adults | 200 ms white noise burst and a 200 ms speech stimulus | Noise CI: 30.8±10.0<br>Speech CI: 29.1±12.7<br>Noise NH: 6.7±1.1 | 75.2±0.2 |

## A.3. Dominant localization cues for BiCI users and NH listeners

Auditory stimuli can be artificially manipulated such that changes in location result only in changes in ILD cues or ITD cues. Studies that manipulate localization cues often present the auditory stimulus by creating a virtual auditory environment for the listener. This is done by measuring the subject’s Head Related Transfer Function (HRTF). Microphones inside the listener’s ear canals pick up sounds that travel from different directions through the subject’s head and pinnae. The HRTF in the frequency domain is calculated by taking the measured sound (output) in the frequency domain and dividing it by the original sound (input) in the frequency domain: $H(f) = \frac{Output(f)}{Input(f)}$

The input/output is generated/measured in the time domain, thus the frequency-domain version is calculated via a Fourier transform. The HRTF in the time domain is calculated via a Fourier transform as well and applied to any other original stimulus in order to introduce the appropriate localizations cues. There is one HRTF for each loudspeaker direction in the virtual auditory environment. The result is delivered to the subject via headphones and the subject perceives the sound to be coming from a specific direction [15,16].

Aronoff et al. (2012) measured the ability of 42 NH listeners to perform a localization task with 12 different target locations spanning a range of +97.5° to +262.5° azimuth. Aronoff et al. (2012) created ILD-only versions of the auditory stimulus by replacing the fine structure of each HRTF function with the fine structure of the HRTF function corresponding to 180° azimuth The substitution of a HRTF fine structure corresponding to one azimuth angle with a HRTF fine structure corresponding to 180° azimuth is equivalent to artificially setting the ITD cue to zero. Thus it is important to be aware that ILD-only cues do contain ITD information however the ITD cue remains constant across azimuth angles. Aronoff et al. (2012) created ITD-only versions of the auditory stimuli by replacing the envelope of each HRTF function with the envelope of the HRTF function corresponding to 180° azimuth. The substitution of a HRTF envelope corresponding to one azimuth angle with a HRTF envelope corresponding to 180° azimuth is equivalent to artificially setting the ILD cue to zero. Thus it is important to be aware that ITD-only cues do contain ILD information however the ILD cue remains constant across azimuth angles. The ILD-only and ITD-only stimulus versions along with a version that contains both ILD and ITD cues together were presented to the listeners during the localization task. Listeners showed similar performance with the ITD-only stimulus as they did with stimuli containing both

ITD and ILD cues; whereas, the ILD-only stimulus resulted in worse performance, thus suggesting that NH listeners rely mostly on ITD cues.

Seeber and Fastl (2008) tested two subjects with BiCIs on a localization task with 11 different target locations spanning the range of -50° to +50° azimuth. Seeber and Fastl (2008) created ITD-only and ILD-only versions of the auditory stimulus by low-pass filtering and high-pass filtering the stimulus, respectively. ILDs are mainly affected by the sound's interaction with the subject's head and are thus restricted to high frequencies that match a wavelength on the order of the head size. ITDs are useful only with low frequencies due to a biological limitation. The auditory nerve has a refractory period that limits the rate at which it fires. The amplitude of the auditory nerve firing reflects the amplitude of the auditory stimulus. Therefore a stimulus with a high rate of amplitude modulation cannot be accurately captured by the auditory nerve. Essentially, two ITDs that differ by an integer number of periods could induce the same firing rate in the auditory nerve when dealing with high frequencies (small periods). The two BiCI users in the study showed similar performance on the ILD-only stimulus as they did with both ITD and ILD cues available, whereas the ITD-only stimulus resulted in much worse performance. This study among others (e.g. Aronoff et al., 2012) argued that ILDs are the dominant cues for localization by BiCI users.

### A.4. Stimulus features influencing localization ability

The level of difficulty during a localization task can be of great interest when designing a study. In a localization training study for example, one would want the localization task to be difficult enough so that subjects have room for improvement but not too difficult so that subjects are able to improve. In addition to the loudspeaker array structure, some stimulus features can influence the difficulty level of a localization task.

Rakerd and Hartmann (1986) tested 4 NH adults on a localization task with 12 loudspeakers spanning the range of -16.5° to +16.5° azimuth. Rakerd and Hartmann (1986) used different stimuli with onset times ranging from 0 ms to 5000 ms and showed that tones with longer onset times correlated with worse localization performance. Moreover, tones with varying durations ranging from 5 ms to 2000 ms were used, showing that the worse localization performance was achieved for durations between 5 ms and 50 ms.

When listeners are asked to localize a single stimulus in a localization task with multiple and simultaneous or temporally proximal stimuli, the other stimuli can mask the target stimulus. Depending on the masker's location, it can increase the difficulty level of the localization task. Lorenzi et al. (1999) tested 4 NH adults on a localization task with 11 loudspeakers spanning the range of -90° to +90° azimuth. Subjects localized a 300 ms click train, composed of 23 $\mu s$ pulses, repeated at a rate of 100 Hz. The masker was a 900 ms white noise stimulus, low pass filtered at 14 kHz and presented at -90°, 0°, or +90° azimuth. Lorenzi et al. (1999) showed that a masker at -90° or +90° azimuth resulted in worse localization performance than a masker at 0° azimuth.

## B. Audio-Visual Integration

It might be the case that a BiCI user would develop a non-ideal auditory map of space due to the degraded localization cues available to them. A non-ideal auditory map of space is a mapping that creates a mismatch between certain spatial locations of auditory sources and the auditory perception of these sources in space. The mismatch is a result of the specific processing of the information available in the stimulus. Early stage processing, by the CI, can decrease the amount of available information that is relayed to the auditory nerve. Later stage processing, by the auditory cortex might not accurately reflect the reliability of each cue unless the information regarding reliability is made accessible to the BiCI user.

The term "plausibility hypothesis" (Rakerd and Hartmann, 1985) refers to the theory that interaural parameters are assessed by listeners to determine their plausibility, given the information listeners have about the environment (e.g. visual images). Thus, in the absence of information about the environment, listeners might weight different interaural localization cues in a way that doesn't necessarily reflect their plausibility. This leaves room for improvement if such listeners are trained with salient cues, and if those cues can be used to provide listeners with stable and consistent perceptual representation of space, thus leading to optimal weighting of auditory localization cues. Even though CI users encounter salient visual cues outside of the laboratory, the natural occurring frequency of those cues might not be high enough to have a significant effect on the reweighting of auditory localization cues [22].

Audio-Visual integration training (AVIT) assumes that the dominance by the visual spatial mapping mechanisms will induce re-mapping or realignment of auditory spatial maps. AVIT relies on a visual stimulus that coincides spatially and temporally with the sound source, serving to reinforce a more fine-tuned auditory spatial map. AVIT has been shown to be an effective tool for altering listeners' auditory spatial maps both in animals and humans [23–25].

Another possible benefit of AVIT lies in the plastic nature of the auditory and visual cortices. The development of a non-ideal auditory map of space could involve plasticity in cortical regions including the auditory cortex and visual cortex, a common occurrence after CI surgery or any significant trauma to the auditory system [26]. One type of such plasticity is cross-modal reorganization of cortex areas [27], specifically audio-visual (AV) reorganization. Such reorganization might be driven by the disparity between the cross-modal cue saliency. Essentially, the superiority of visual information over auditory information for a CI user may cause auditory information to be considered unreliable compared to visual information. Consequentially, cortical regions normally used for auditory processing may be reassigned for visual processing [27]. Thus, when designing a training method aiming at imparting a better auditory map of space for CI users, it may be beneficial to utilize AV cues in order to reverse such reorganization, especially when considering CI users' propensity for AV integration as discussed below [28].

### B.1. The use of AVIT for sound localization

Vision plays a crucial role in aligning neural representation of space in the brain, and is used to resolve spatial conflicts between modalities like audition and vision. This role has been demonstrated in many studies and therefore the ability to integrate auditory and visual information is critical for the outcome of CI user rehabilitation [29].

Visual stimuli have been shown to influence the spatial perception of auditory stimuli after prolonged duration of AV integration in both humans and animals. Knudsen and Knudsen (1989) reared 12 barn owls that were fitted with binocular prisms at birth. The prisms displaced the visual field to the right by 11°, 23°, or 34°. The barn owls performed a localization task by turning their head towards auditory and visual stimuli originating from a moveable loudspeaker and photodiode. The head orientation of the owls was recorded for both visual and auditory stimuli and the difference between those orientations was used to quantify the effect of the prism. The audio-visual difference was recorded just before the prisms were taken off (64 to 202 days after prisms were put on originally). The results showed no significant audio-visual difference for the 11° and 23° prisms, indicating that the auditory spatial map for the owls was shifted to the right by 11° and 23° respectively. The 34° prism induced between 53% to 66% shift in the owls' auditory spatial maps, suggesting that there is a limit to the influence of AVIT over the auditory spatial map of owls.

A similar study was done with nine NH human subjects by Zwiers et al. (2003). Subjects wore binocular lenses that both covered the visual field beyond a 20° radius and compressed the available visual field by half. The lenses were worn for two to three days with the exception of sleeping hours, allowing the subjects to integrate auditory and visual stimuli during their normal day and during controlled conditioning sessions. The subject performed a localization task before and after the aforementioned adaptation period. It included a 150 ms Gaussian white noise auditory stimulus and 87 possible target locations spanning -50° to +50° in azimuth and -22° to +22° in elevation. The shift in auditory spatial map manifested most prominently in the azimuth, where the spatial auditory field was increasingly compressed away from the center in the available visual area, and plateauing at a compression of 5° in the visually covered area.

Recanzone (1998) tested 3 NH adults on a localization task with incongruent visual and auditory stimuli. Subjects localized 200 ms tones (750 Hz and 3000 Hz) that could originate from 15 different locations spanning the range of -28° to +28° azimuth. The subjects were trained with a visual stimulus in the form of an LED flash coming from 8° to right of the auditory stimulus. The training procedure lasted 20 to 30 minutes. Subjects showed an 8° shift in auditory perception when tested with only the auditory stimulus after the training period.

Strelnikov et al. (2011) plugged a single ear for each of the 18 NH adult subjects and tested them on a monaural localization task. Six subjects were trained with visual and auditory stimuli consisted of LED flashes and 50 ms white noise bursts originating from 15 possible locations spanning the range of -70° to +70° azimuth. Six other subjects were similarly trained but with only an auditory stimulus. The six remaining subjects were trained with an auditory stimulus and a post selection behavioral feedback in the form of a visual sign indicating 'correct' or 'incorrect' response. The auditory only group showed minimal improvement in localization error (0.9° ± 0.1°) whereas both the AVIT group and the behavioral feedback group showed large improvements (13.6° ± 0.1° and 10.8° ± 0.1° respectively). Thus, AVIT proved to be not only an effective method to improve localization ability in a sub optimal listening environment, but also more successful than other training methods like behavioral feedback.

### B.2. Echo suppression via audio-visual integration

When localizing a single sound source, audio-visual integration can help to promote weighting of the localization cues associated with a particular sound in an optimal way. A similar situation occurs when a listener is asked to localize a single sound among multiple available sounds, especially when the sounds are presented simultaneously or temporally proximal. A visual cue that is spatially and temporally coincident with the desired sound can help weight the localization cues associated with it more heavily than those associated with the other sounds, thus allowing the desired sound to be localized more easily.

Bishop et al. (2011) tested 45 NH adults on a discrimination task. Two temporally proximal sounds (delayed by several milliseconds) were presented from two separate locations at either -18° or +18° azimuth. Subjects were asked to indicate whether they heard two sounds or one sound (The phenomenon of echo suppression is called the precedence effect and is described in section C). The listeners performed this task in three different conditions. The first condition consisted of a visual stimulus (LED) coinciding with the leading sound ($AV_{lead}$). In the second condition the visual stimulus coincided with the location of the lagging sound ($AV_{lag}$), and in the control condition no visual stimulus was present. The results showed significantly more reports of two sounds for the $AV_{lag}$ condition compared to control and significantly less reports in the $AV_{lead}$ condition compared to control, suggesting that a visual stimulus can diminish or enhance echo suppression.

### B.3. CI users are superior AV integrators

CI users have a higher propensity for AV integration due to their higher ecological need to supplement heard speech with lip reading compared to NH individuals. Moreover, Rouger et al. (2007) showed that CI users are better at integrating auditory and visual cues compared to NH individuals by testing 97 CI users and 163 NH subjects on a speech identification task in two conditions: auditory alone condition (*A*) and an audio-visual condition (*AV*) where the visual stimulus was a video of the person producing the auditory speech stimulus (focused on their face). The subjects' performance on the speech identification task was measured in terms of the percentage of word identified correctly. In order to avoid ceiling effects for the NH listeners, the speech stimulus was degraded either by adding a white noise masking stimulus or by vocoding the speech stimulus (see section C.3 for more information on vocoding). The benefit from integrating both the visual and auditory stimuli was calculated by taking the normalized difference between the subject's results on each condition: $Benefit = \frac{AV-A}{100-A}$

The results showed a significantly higher AV benefit for CI users than for NH subjects, which points to a possible advantage of tailoring rehabilitation tools for CI users around audio-visual integration.

### C. Sub-optimal listening environments and the precedence effect

The aim of this study is to develop a training method that could be used to improve the localization ability of CI users. It would be informative to test the efficacy of the AVIT method on NH listeners before transitioning to a study involving CI users, as less confounding factors are associated with NH listeners compared to CI users. There are many factors affecting CI users' performance, including electrode array placement, the degree of spiral ganglion cell survival, cause of deafness, duration of deafness prior to implantation, age at implantation, delay between implantation of the first CI and the second CI, number of months post CI activation at time of study, number of electrodes in CI, processing strategy in CI, amount of neural degeneration due to auditory deprivation, age and other factors that are not well understood [12,13]. These factors could potentially increase the variability in the results of this study, thus making it more difficult to attribute improvement in localization ability to the training paradigm rather than one or more of the aforementioned factors.

NH listeners make smaller localization errors compared to CI users (6.7° vs. 30.8° RMS error respectively according to Grantham et al., 2007). The small localization errors made by NH listeners provide little room for benefit from AVIT. However, it is possible to degrade the performance of NH listeners by creating a listening environment that would simulate some of the difficulties faced by CI users. In such an environment, NH listeners would make large localization errors, thereby allowing testing the efficacy of the AVIT method. Two such environments were created. One environment where echoes are introduced into the sound localization task [33] and one where CI listening conditions are simulated by vocoding the auditory stimulus [34]. When testing NH listeners in an echoic environment, LEDs in the AVIT paradigm would signal the location of the first arriving sound, thereby allowing the listener to weight its location more heavily than that of the later arriving sounds (echoes). Sound localization in the presence of echoes is an ecologically relevant problem. Moreover, testing the AVIT paradigm in this scenario should indicate whether AVIT could affect the way different auditory cues are weighted and interpreted. This effect may be important when fine-tuning the auditory spatial map in situations relevant to CI users for whom some auditory localization cues are more reliable than others [17].

In order to understand the parallels between the weighting of auditory cues in an echoic environment and the weighting of auditory cues by a CI user, one must understand a phenomenon called The Precedence Effect (PE). An echoic environment can be challenging for a NH listener depending on the echo's temporal and spatial proximity to the original signal. The PE refers to the phenomenon of echo suppression, whereby the ability of listeners to localize the first arriving sound is facilitated. However, the extent of echo suppression, and more generally the clarity and the location at which the two sounds are perceived depend on the specific temporal delay between the two sounds.

*Summing localization* refers to a case where a short delay (0 to 1 ms) results in an integrated singular image of the first and second sound (from now on referred to as *lead* and *lag*). The simplest case of summing localization occurs when there is no temporal overlap between the lead and lag. In this case the perceived location is a spatial average of the lead and lag's locations. In cases where an overlap does occur, a more complex average take place that takes the lead and lag's amplitudes and phases into account [35].

*Fusion* refers to a state in which a medium delay (1 to 5 ms) results in a perceptually fused singular image whose location is dominated by the leading sound. Other nomenclatures for this case include *law of the first wave front* and *localization dominance* [36]. Finally, *lag* *discrimination suppression* refers to the case in which a long delay (> 5 ms) results in the lag's stimulus parameters becoming less discriminable due to the presence of the lead [37]. As the delay get larger the lead's dominance over the lag begins to decrease until the lead and lag are perceived as two separate auditory events.

The Echo Threshold (ET) is sometimes defined as the limit between two fused auditory events and two separate ones [35]. This term has been used in the literature with different definitions corresponding to different tasks that were used to measure this threshold. For example, Freyman et al. (1991) tested subjects on a task in which listeners indicated whether they detected the lagging sound, whereas, Yang and Grantham (1997) had participants identify the loudspeaker that generated the lagging sound. The former involved a subjective response by the listener whereas the latter involved a response for which accuracy can be measured.

For delays larger than the echo threshold delay, the temporal order of the two stimuli becomes ambiguous. Moreover, some subjects report to perceive the lead's location to be as shifted towards the actual location of the lag and vice versa. However, other subjects might perceive the location of both the lead and lag to originate from the actual location of the lead. These factors of stimuli temporal order and separability can generate spatial errors when a subject tries to localize either the leading or lagging sound at or above the echo threshold. In the present study listeners are trained to localize the lead for a delay at their ET. This type of learning is made possible by weighting the spatial cues associated with the lead more heavily than the spatial cues associated with the lag, which is essentially what happens during fusion. In other words, the listeners' ETs are raised.

### C.1. Buildup and breakdown of the precedence effect

As mentioned above, it is possible to change the specific value of a listener's ET. In this study AVIT is used to do that. Testing the efficacy of the AVIT method would require that the ET be not altered due to other factors. One such possible factor is called the *buildup* of the PE.

Tolnai et al. (2014) studied the PE in three ferrets and four humans. The findings of this study showed that if the locations of the lead and lag as well as the lead-lag delay are constant over several stimuli presentations, then subjects tend to localize the leading sound more accurately than if either the locations of the lead and lag switch or the lead-lag delay changes randomly. This suggests that under the aforementioned constant conditions the ET is raised so that the perception of the lead dominates that of the lag. The study also showed that once the buildup effect occurs and the ET is raised, it is possible to revert to the original state by switching the locations of the lead and lag. This is called the *breakdown* of the PE.

Freyman et al. (1991) presented similar buildup results by testing 4 NH adults who were asked to indicate whether or not they heard the lagging sound. The subjects showed a lower percentage of lag reports when the lead and lag locations remained constant.

In the present study, the effect of AVIT on localization at the ET is of great interest. Therefore, the locations of the leading and lagging sounds are randomly changed during the experiment.

### C.2. The effect of spatial separation on echo threshold

Litovsky and Colburn (1997) suggested that spatial separation between the lead and lag significantly reduces the magnitude of the ET delay. In contrast, Yang and Grantham (1997b) claimed that there was no consistent effect of the spatial separation between the lead and lag sources on echo suppression. Six NH adults indicated whether they heard one sound or two sounds, for different lead-lag delays and for different lead-lag spatial separations. For each lead-lag spatial separation, the lead-lag delay was adjusted adaptively [43], increasing for a “one sound” response and decreasing for a “two sounds” response in order to find the echo threshold. As the spatial separation increased, some subjects showed an increase in ET while others either showed a decrease in ET or showed no change in ET.

In the present study the ET was measured in a similar adaptive process as described above. The spatial separation when the ET was measured was small (20° azimuth). This ensures that when listeners are trained with AV integration to better localize the leading sound, they are either doing so at their ET (for small spatial separation) or at a delay slightly above their ET (for larger spatial separations). In both cases the task of localizing the leading sound is still challenging enough to allow for improvement via AVIT.

### C.3. Vocoding

Vocoding refers to the processing of an auditory signal, usually in a way that simulates the way a cochlear implant would process the signal. Vocoding has been used to investigate the difficulties CI users encounter by testing the performance of NH listeners with vocoded stimuli [44,45]. The vocoding process consists of three stages. First, the signal is digitally band-passed filtered into a number of frequency channels. Second, the amplitude envelope in the time domain of each channel is extracted by half-wave rectifying and low pass filtering the signal in each channel. Half-wave rectification is equivalent to taking the absolute value of the signal in the time domain and thus mediating the high frequency influence of the signal’s fine structure on the envelope shape. Finally, the envelope in each channel is used to modulate a carrier that replaces the original fine structure of the signal and the channels are summed together (see Figure 1). Common carrier types are sine tone, white noise, and Gaussian envelope tone (GET) carriers. The sine tone and white noise vocoder have the advantage of preserving the signal's envelope. The white noise carrier has the advantage of having a "fuller" spectrum, whereas the sine tone carrier has the advantage of having a deterministic spectrum. The GET vocoder (see Figure 2) is composed of pulse trains that are scaled by the signal's amplitude in each channel. It uses a Gaussian envelope to modulate the center frequency of each channel. Each pulse train has the same pulse rate. One can argue that the GET vocoder simulates CI processing most reliably as it uses periodic pulse trains and a single pulse rate similarly to a CI processing scheme. In this study a noise carrier is used because its non-deterministic nature reduces the chance of NH listeners having access to an ITD localization cue that would not be available to a CI user.

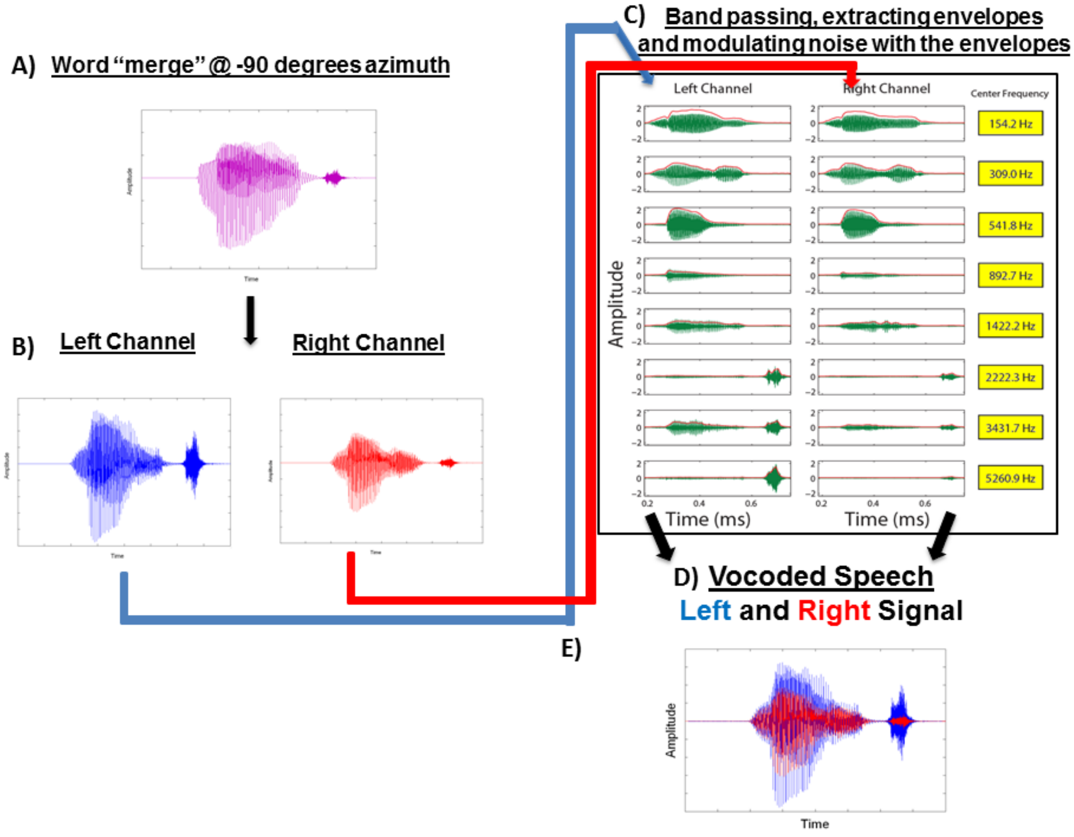


Figure 1. Jones et al 2012. Vocoding process schematic.

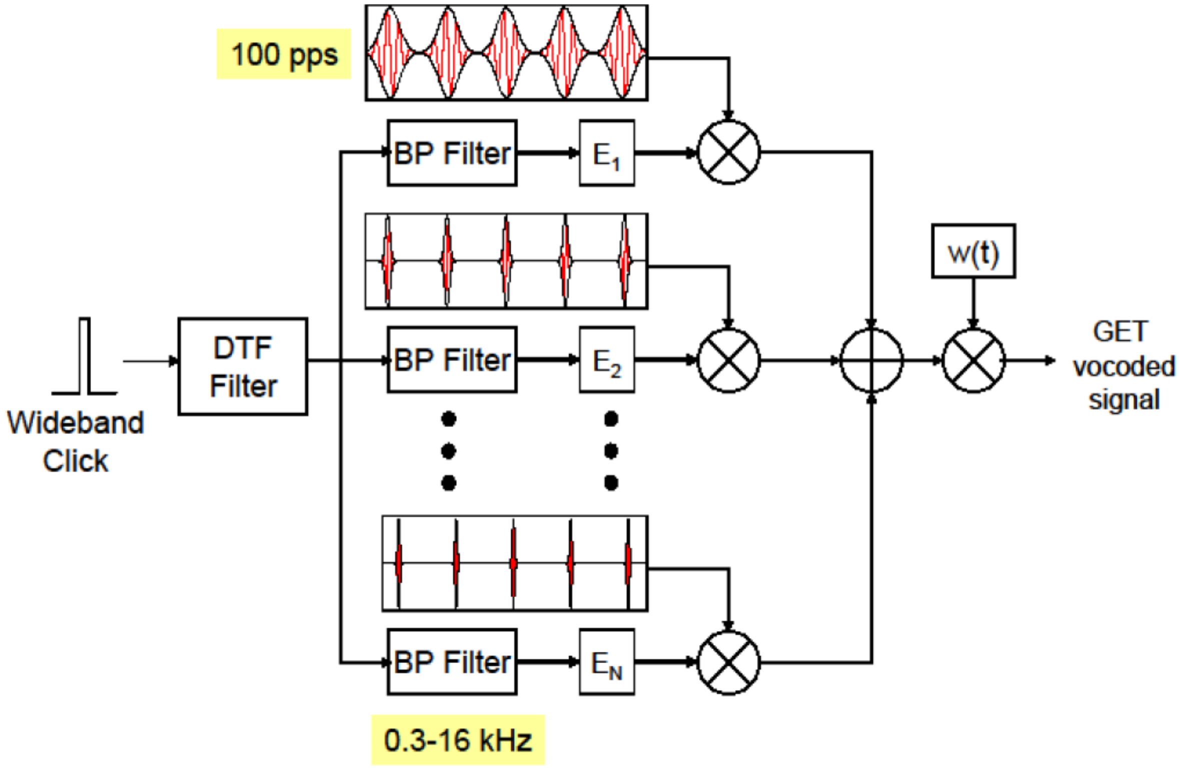


**Figure 2.** Goupell et al (2010)**. Gaussian Envelope Tone (GET) vocoder schematic. The signal is (a) filtered through a Directional Transfer Function (DTF), also known as Head Related Transfer Function (HRTF), (b) band pass filtered where the energy in each channel scales a train of Gaussian envelopes the modulates a sine tone with the channel's center frequency, (c) summed across channels, and (d) ran through a temporal window with on and off ramps. The pulse rate is the same for every channel (100 pulses per second).**

## D. Training and learning

Auditory spatial perception can be modified [23,24,31]. However, the amount and frequency of training needed to learn and retain a new auditory spatial map is task dependent [22]. Other learning determinants include task difficulty [47], central factors like attention and motivation [48], similarity between testing and training procedures, and whether the task is realistic or controlled [49]. For example, a speech perception task that involves an alternating number of maskers is more realistic than one that involves the same number of maskers. It has been shown that training regimens that are more realistic (i.e. have more complexity) are more likely to generate learning beyond the trained stimulus [50]. The ability to transfer learning to new tasks and/or new stimuli than the one trained on is called *generalization*. A more realistic localization-training paradigm that includes visual cues is hypothesized to be more likely to facilitate generalization.

The specific type of learning is often of interest when discussing training studies. Perceptual learning describes the long-lasting improved sensitivity to a stimulus that occurs due to extensive exposure to that same stimulus [51]. The encoding process of a stimulus in the periphery and certain higher cortical areas can be modeled by a signal plus internal and external noise [52]. It is the suppression of internal and external noise that is thought to explain perceptual learning [53]. Gold and Watanabe (2010) suggested that unlike perceptual learning, other forms of learning might establish task rules, associations and strategies. However, Goldstone (1998) claimed that delineating human learning into perceptual and other types of learning is regrettable as it causes fruitful links between them to be neglected. One of the aspects of perceptual learning described by Goldstone involves the weighting of a stimulus' relevant dimensions compared to its dimensions that are irrelevant to the task at hand (e.g. salient localization cues vs. degraded localization cues). Although this aspect relates to the type of learning hypothesize to occur in the present study, AVIT cannot be claimed to invoke only perceptual learning as other types of learning may occur as well. The more important distinction to be made for this study is between training mechanisms driven by feedback or reinforcers (supervised learning) that can trigger different types of learning and ones without feedback (unsupervised learning) that can only rely on natural neural tendencies to specialize for certain stimulus features by suppressing internal and external noise in the environmentally supplied stimuli.

### D.1. Amount of training needed for learning

One of the main unresolved questions with training studies is how much training is needed to induce a specific type of learning called *consolidation*. Consolidation refers to the transfer of training-induced learning from short-term memory to long-term memory [55].

Wright and Sabin (2007) tested 28 NH adults that were split into 4 groups. Two of the groups performed a frequency discrimination task while the other two performed a temporal-interval discrimination task. For each task, one group was trained with 360 daily trials and the other with 900 daily training trials. Training took place over 6 separate, usually consecutive days. The respective discrimination threshold was measured in each day. Subjects performing the frequency discrimination task first listened to two tones with identical frequency $(f)$ separated by a time interval $(T)$ and then listened to two tones separated by the same time interval $(T)$ with identical frequency but shifted compared to the first two tones $(f \pm \Delta f)$. Subjects were asked to indicate which of the tone pairs had a lower frequency. Subjects performing the temporal-interval discrimination task listened to two tones with identical frequency $(f)$ separated by a time interval $(T)$ and then listened to two tones with the same frequency $(f)$ but a time interval that is shifted compared to the first two tones $(T \pm \Delta T)$. Subjects were asked to indicate which of the tone pairs had a longer time interval. The results showed that both the short and long daily training groups significantly improved (had lower thresholds) on the temporal-interval discrimination task while only the long daily training group improved on the frequency discrimination task. After adjusting for the difference in total training trials, Wright and Sabin concluded that there exist a minimal number of daily training trials needed for consolidation and that number is task dependent.

In contrast to the conclusion above, Nogaki et al. (2007) showed that the frequency of training (number of trials per unit of time) is not significant for consolidation. Eighteen NH listeners were trained on vowel recognition of vocoded speech. The 18 subjects were divided into 3 groups. One group was trained once per week, another group was trained 3 times per week and the last group was trained 5 times per week. Each training session lasted 1 hour. The results showed comparable improvement in vowel recognition across training groups. Nogaki et al concluded that the outcome of auditory training might depend more strongly on the total number of training trials than the frequency of training.

### D.2. The effect of motivation and attention on learning

Bergan et al. (2005) investigated the role of motivation and attention in the acquisition of new auditory maps of space. Twelve owls were fitted with prisms that displaced the visual field by 17° horizontally. The owls were divided into two groups. One group was fed dead mice while the other group hunted live mice and therefor was associated with higher levels of attention and motivation. Hunting and feeding lasted 1 hour in each day. The effect of integrating auditory stimuli and shifted visual stimuli on the auditory map of the owls was assessed by measuring the ITD tuning of layers in the owls' superior colliculus. The superior colliculus produced gaze shifts (head and eye movements) in response to a stimulus and is therefore selectively sensitive to ITD information in the stimulus that changes as a function of azimuth angle. The results showed that only the hunting group exhibited a consistent and significant shift in ITD tuning (25 $\mu s$ which is equivalent to 10° azimuth), suggesting that motivation and attention play a crucial role in learning induced by AVIT.

### D.3 Scaffolding: A gradual adjustment of task difficulty

Linkenhoker and Knudsen (2002) performed a similar experiment to the one described in the previous section but with older owls that were shown to be resistant to changes in their auditory map of space even when allowed hunting their food. These owls showed only a 4 $\mu s$ change in ITD tuning (equivalent to 1.6° azimuth) for 23° prisms after a few months. Furthermore, a group of 5 owls was tested with prisms of adjustable strength. This group started out with 6° prisms for 21 days and exhibited a 15 $\mu s$ shift in ITD tuning (equivalent to 6°). The prism strength was then adjusted to induce an 11° visual shift which resulted in a 20 $\mu s$ shift in ITD tuning (equivalent to 8°). Finally the prism strength was then adjusted to induce a 17° visual shift which resulted in a 25 $\mu s$ shift in ITD tuning (equivalent to 10°). These results point to the power of training schemes that employ tasks with gradual increments in difficulty level.

### D.4. Generalization and task complexity

There have been many training studies showing the failure of subjects to generalize learning. For example, Rieser et al. (1995) trained 8 adults by having them run on a treadmill while being towed by a vehicle that was traveling at a speed higher or lower than that of the treadmill. Subjects were then asked to visually estimate their distance to a target and walk towards that target while blindfolded. The results showed that subjects who were trained with a higher vehicle speed (compared to the treadmill's speed) undershot the target and those with a lower vehicle speed overshot the target. This effect did not generalize to other tasks like throwing or turning in place however it did generalize to sidestepping. One could argue that sidestepping is much more similar to the trained task as both consist of changing one's location without changing the facing direction. Perhaps a more complex training scheme like towing the treadmill in a circular path would have induced more generalization.

Logan et al. (1991) trained 6 native speakers of Japanese on a speech discrimination task that included pairs of English words contrasting /r/ and /l/ (e.g. Rock vs. Lock). The training scheme specifically contained words spoken by multiple talkers in order to increase task complexity. Unlike previous studies that showed improvement on the trained words but not on novel words, the Logan et al study showed improvements on trained words, novel words spoken by a familiar talker, and on novel words spoken by a novel talker.

These studies suggest that a training scheme is more likely to induce learning on a task that is similar to the training task. Moreover, real-world tasks often include a high level of complexity and thus increasing the level of complexity in a training scheme could potentially induce generalization to real-world scenarios.

Chapter 1 informed the reader of the difficulties faced by BiCI users when localizing sounds, specifically regarding the mismatch between salient and degraded localization cues. AVIT was introduced as a method that has been used to alter auditory spatial maps, in some cases doing so by allowing the listener to weight certain localization cues more heavily than other localization cues. In this method, a visual stimulus is used to reinforce the association between the former localization cues and the desired location in space. Two sub-optimal listening environments were discussed in the context of simulating a mismatch between localization cues. Chapter 2 describes the echoic environment in which listeners are instructed to localize a leading soundwave in the presence of other lagging soundwaves. AVIT is used to allow the listener to weight the cues associated with the leading soundwave more heavily that those associated with the lagging soundwaves. Chapter 3 describes the CI simulated environment, in which listeners are asked to localize a single vocoded stimulus that contains salient ILD cues and degraded ITD and spectral cues. AVIT is used to allow the listener to weight the salient cues more heavily than the degraded ones.

# II. EXPERIMENT 1: LEAD LOCALIZATION AT ECHO THRESHOLD CONDITIONS

Experiment 1 aims to test the hypothesis that AVIT can allow listeners to weight certain localization cues more heavily than other localization cues, thereby improving the listeners' localization ability.

### A. Listeners and equipment

Twelve NH adult subjects (mean age 22, range 18-31) participated in the study. They were divided into a visually-trained group and a control group (six subjects in each group). All subjects reported no auditory or neurological disease and had normal or corrected to normal vision. As verified by hearing screening, the listeners had pure tone thresholds within 20 dB of audiometrically normal hearing at octave interval frequencies between 250 and 8000 Hz, and the thresholds between the two ears differed by less than 15 dB at any tested frequency.

The experiment was held in a dark sound booth padded with sound absorbent material. Subjects sat on a chair in front of a touch screen monitor, surrounded by semi-circular arrays of loudspeakers and light emitting diodes (LEDs). The apparatus consisted of 19 loudspeakers equally spaced around the range of -90° azimuth to +90° azimuth and was covered by a sound transparent black cloth. The LED array spanned the same azimuth range with each LED situated 7'' (18 cm) below each loudspeaker. After being presented with the auditory stimulus, subjects indicated the source of stimulation by operating a graphic user interface displayed by the touch screen monitor (Figure 3). Hardware included a Tucker-Davis System III (Tucker-Davis Technologies, Alachua FL). Customized software for stimulus presentation and data collection was written in Matlab.

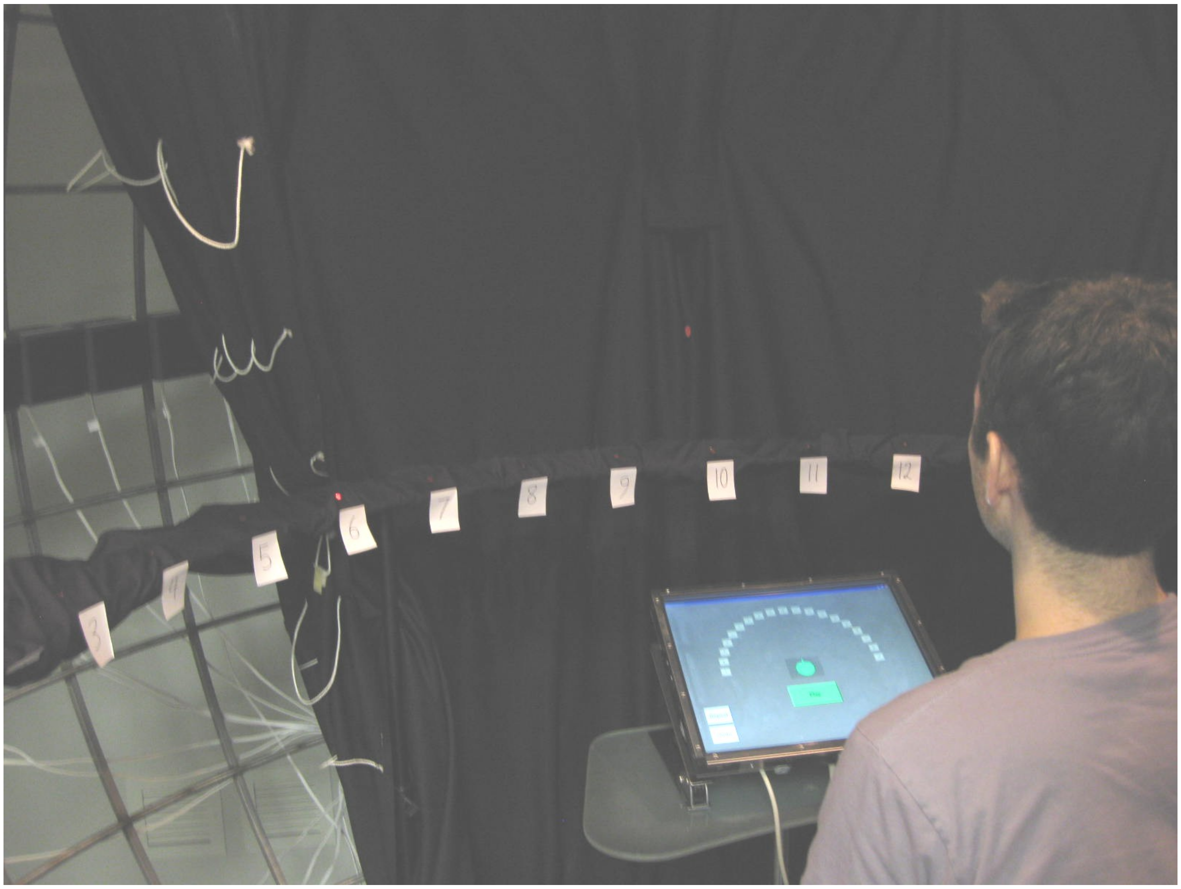


Figure 3. Photo of the loudspeaker array (with black cloth pulled open), LED array and graphic user interface used by the listeners.

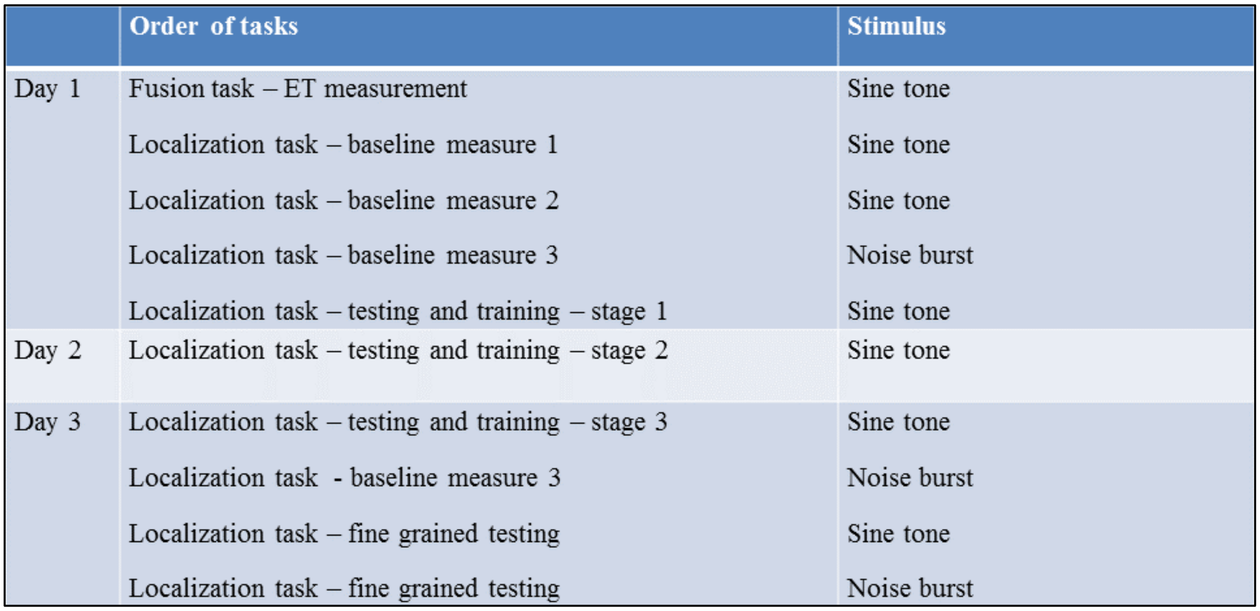
## B. Stimuli

Stimuli consisted of a 60 dB, 40 ms, 2000 Hz sine tone turned on at positive-going-zero crossings with no ramp, and a 40 ms white noise burst with no ramp. Subjects received training with the former stimulus but not with the latter in order to ascertain the potential of AVIT to generalize to an untrained stimulus. The need for a difficult (but not too difficult) localization task dictated the choice of auditory stimuli. Rakerd and Hartmann (1986) showed that tone pulse durations between 5-50 ms induced high localization errors both in single source and echoic environments. Moreover, Rakerd and Hartmann showed that tones with longer ramp durations induced localization errors that varied in size depending on target location. Conversely, tones with no ramps resulted in smaller localization errors that were of similar size across target locations. Sound level measurements in the lab's sound booth showed that level cues (due to reflections off padded walls) regarding azimuthal location are not significant above 2000 Hz, which dictated the choice of frequency.

### C. Procedure

Subject participation in the study took place on 3 separate days, all within a 7-day time span. On each day, participants were tested and trained on a localization task. The number of loudspeakers in the task increased from day to day in order to promote a gradual training scenario. Subjects in the visually-trained group received a reinforcing visual stimulus during training. This visual stimulus was absent during the training stage of the control group. The subjects' localization ability was measured pre- and post-training in order to assess improvement. Wright and Sabin (2007) showed that there exists a minimal amount of training trials per day needed for consolidation and that this amount is task-dependent. Therefore, the amount of training in this experiment was influenced by a previous AVIT study that was successful in improving localization ability in sub-optimal listening conditions [31]. Specifically, the number of daily training trials in each stage was equal to or larger than the number of daily training trials in the Strelnikov et al. (2011) study. The order of testing in the 3 different study days is summarized in Table 2. The hypothesis was that visually-trained subjects would show statistically significant improvement in localization ability compared to subjects in the control group.

Table 2. Order of testing for experiment 1



| | Order of tasks | Stimulus |
|---|---|---|
| Day 1 | Fusion task – ET measurement | Sine tone |
| | Localization task – baseline measure 1 | Sine tone |
| | Localization task – baseline measure 2 | Sine tone |
| | Localization task – baseline measure 3 | Noise burst |
| | Localization task – testing and training – stage 1 | Sine tone |
| Day 2 | Localization task – testing and training – stage 2 | Sine tone |
| Day 3 | Localization task – testing and training – stage 3 | Sine tone |
| | Localization task - baseline measure 3 | Noise burst |
| | Localization task – fine grained testing | Sine tone |
| | Localization task – fine grained testing | Noise burst |

## Fusion task

In the fusion task listeners were presented with a pair of lead-lag stimuli after which they indicated the lateral position (left/right) of the lagging sound. The stimulus used for the fusion task was the sine tone stimulus. The lead and lag randomly originated from +10° and -10° azimuth, one from each loudspeaker. At short delays below a listener's echo threshold the lead would dominate, making the task difficult whereas at long delays above the ET the lag would be much easier to locate. The ET was estimated by fitting a 3-up-1-down (79% threshold) psychometric function to the data [43]. This method employs regression analysis with the lead-lag delay as an independent variable and the probability of response as the dependent variable. In this procedure the lead-lag delay is varied adaptively depending on the listener's response, increasing the delay each time the subject provides 3 consecutive incorrect responses and decreasing the delay each time the subject provides one correct answer. The amount by which the delay increases or decreases is halved at each “reversal” (the first delay increase following a delay decrease or vice versa). Given enough trials, the delay would asymptotically reach the listener’s ET. In such a case, there would be an equal probability for a delay increase as for a delay decrease. Therefore, the probability for an incorrect answer is $0.79$ ($0.79^3 = 0.5$). However, it is not feasible to test a subject for an infinite number of trials, as the subject’s attention level cannot be maintained for an infinite duration of time. If a subject’s ET measures showed large variability in the first 3 adaptive tracks, two more adaptive tracks were added, in order to gain a better estimation of the subject’s ET. Therefore, each subject performed the discrimination task 3-5 times. The final ET estimation was taken as the average of 3-5 measures from all the adaptive tracks. In order to ameliorate the uncertain nature of the ET estimate, the subject was asked to indicate the number of sounds they heard for several trials with the lead-lag delay corresponding to their ET estimate. If the subject reported hearing one sound only, a perception that corresponds to delays shorter that the ET, the delay was increased (up to 2 ms above calculated ET). Adaptive track samples from two subjects SVS and SVT can be seen in Figure 4.

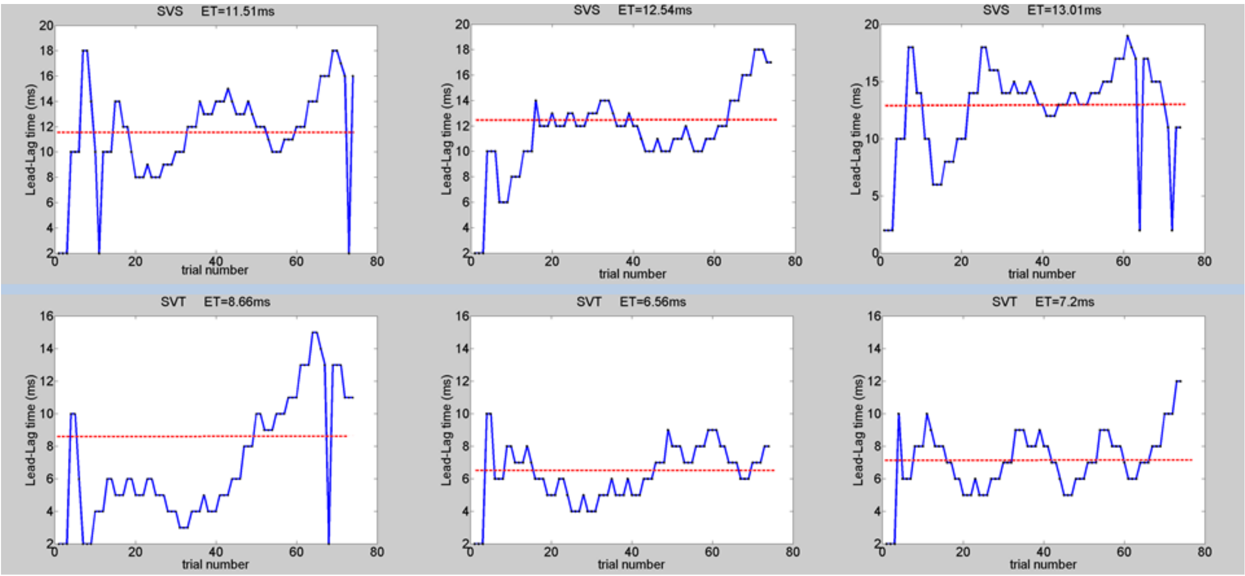


**Figure 4. Adaptive tracks for fusion task. Top row: Subject SVS. Bottom row: Subject SVT. Red line indicating echo threshold. Large dips indicate probe trials that gauged the subject's attention. On these trials subjects hear only the leading sound and thus should have a predictable response. If the subject fails to produce this response (i.e. due to lack of attention) then the testing run is stopped and repeated.**

## Localization task

During the localization task a leading sound and two lagging sounds were presented to the listener. All three sounds were presented at the same intensity. Both lagging sounds were delayed by the subject's echo threshold compared to the leading sound. Subjects were instructed to "locate the position of the first sound". For an N-loudspeaker task the lead would randomly originate from any of the N loudspeakers. The first lagging sound would randomly originate from any of the N loudspeakers except for the loudspeaker from which the leading sound originated on that particular trial. The second lagging sound randomly originated from {-80°, 0°, +80°}. For example, the leading sound could originate from 10°, the first lagging sound from -30° and the second lagging sound from -80°.  Under these kind of echoic conditions, NH listeners may display an error pattern marked by precision errors, similarly to that of CI users under normal conditions [1].

All 3 baseline measurements corresponded to a 6-loudspeaker configuration (-50°,-30°,-10°,+10°,+30°,+50° azimuth). The first baseline measure consisted of 90 trials with the sine tone stimulus as a single source. The second baseline measure consisted of 90 trials with 3 identical sine tones, where the two lagging tones were delayed by the subject's ET compared to the leading tone. The third baseline measure consisted of 90 trials with 3 identical white noise bursts, where the two lagging bursts were delayed by the subject's ET compared to the leading burst. This echoic tone/burst configuration shall be denoted as the *ET condition*. In these baseline measures subjects were instructed to indicate the position of the first sound.

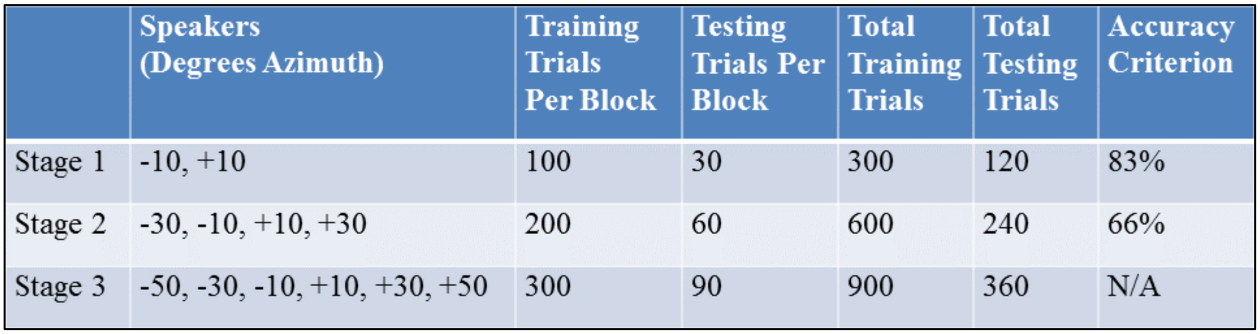
The performance on the second and third baseline measures will be compared to performance on identical measures post-training in order to calculate the amount of localization improvement. All subjects were required to maintain a RMS error greater than or equal to 25° during these baseline measures in order to ensure significant room for improvement. Pilot testing indicated that a listener who improves their ability to a high degree typically improves by about 80%. It was apparent that the best localization ability achievable under these conditions corresponds to about 5° RMS error. Thus, it was determined that the starting point should be at least 25° RMS error in order to allow for 80% improvement. Ideally, the listener would completely suppress the lagging sounds, and only perceive the leading sound post-training. Therefore, a listener's performance on the first baseline measure, where only a single sound is presented, indicates the highest level of performance that can be induced post-training.

Testing and training were performed gradually with one stage for each day (see Table 2). Subjects in the visually-trained group were trained with a one second LED flash, originating from the location of and simultaneously with the leading sound. The subjects were instructed to indicate the position of the light while still trying to locate the leading sound. Three training blocks were alternated with four testing blocks (in which no visual stimulus was present) in order to assess the subject's improvement from one training block to the next. Data detailing the progress of each listener throughout the study procedure can be found in appendix C.

Only subjects in the visually-trained group were required to perform to a predetermined level in order to advance from one day to the next. Details of the training stages are shown in Table 3. Pilot testing indicated the highest level of performance in stage 1 while still allowing for an error due to a rare lapse in attention (83% correct identification of the targets). The percent correct criteria in each stage were set to maintain the same d' sensitivity as that corresponding to the 83% criterion in stage 1, while taking into account the difference in difficulty between the tasks in the different stages [59].

The last testing block in stage 3 was used to compare localization performance on the trained stimulus (sine tone) post-training to results obtained in the pre-training phase. The same task was then performed for the noise burst stimulus. The last localization task in the experiment, denoted *fine grained testing*, consisted of 180 trials in a 12-loudspeaker configuration (-60°,-50°,-40°,-30°,-20°,-10°,+10°,+20°,+30°,+40°,+50°,+60° azimuth; Leading sound could not originate from 0°). It was performed first for the sine tone stimulus and then for the noise burst stimulus. During the *fine grained testing* task subjects localize sounds originating from locations for which they received training and for which they did not. Comparing the subjects' performance on these two types of source locations will shed light on the type of effect the training had on their auditory maps of space. Specifically, the *fine grained testing* task will allow to determine whether subjects merely memorized the relationship between the stimulus and the trained locations, or if their auditory maps of space were altered in a continuous way.

Table 3. Setup of training stages in experiment 1.



| | Speakers (Degrees Azimuth) | Training Trials Per Block | Testing Trials Per Block | Total Training Trials | Total Testing Trials | Accuracy Criterion |
|---|---|---|---|---|---|---|
| Stage 1 | -10, +10 | 100 | 30 | 300 | 120 | 83% |
| Stage 2 | -30, -10, +10, +30 | 200 | 60 | 600 | 240 | 66% |
| Stage 3 | -50, -30, -10, +10, +30, +50 | 300 | 90 | 900 | 360 | N/A |

## D. Results

When analyzing the data, the amount of improvement in localization ability was estimated by looking at two measures: *Error Improvement* and *Normalized Error Improvement*, which compensates for the variability in the subjects' initial pre-training error.

$$Error\ improvement \equiv EI = E^{Pre} - E^{Post}$$

And

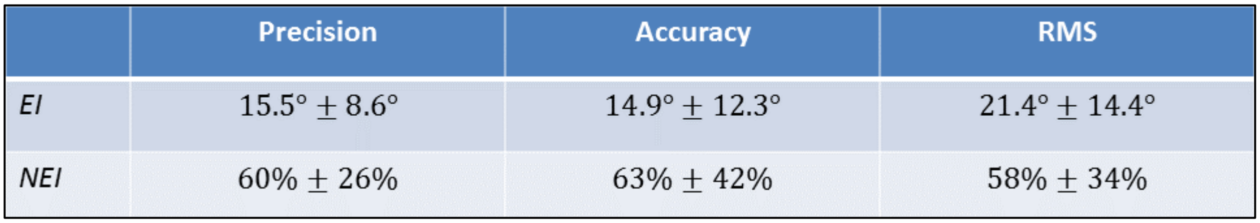
$$Normalized\ error\ Improvement \equiv NEI = \frac{E^{Pre} - E^{Post}}{E^{Pre}} \times 100\%$$

Where $E^{Pre}$ is the error prior to the first training session and $E^{Post}$ is the error after the last training session. Three related concepts were of interest in this experiment: dominant error types, cause of improvement, and generalization to untrained conditions.

### D.1. Dominant error types

The pre-training precision errors of all subjects (visually-trained and control groups) were compared with the pre-training accuracy errors of all subjects via a Kruskal-Wallis test. No statistically significant difference was found both for the tone and noise stimuli (H(1,22) = 0, p = 1 and H(1,22) = 1.02, p = 0.3122 respectively), suggesting that precision and accuracy errors affected listeners equally on average. Nonetheless, some subjects were clearly plagued by one error type more than the other. Subject SWB exhibited better accuracy than precision (Figure 5) while subject SWQ exhibited better precision than accuracy (Figure 6). The comparable improvement measures on precision and accuracy exhibited by the visually-trained group for the trained stimulus can be seen in Table 4 along with the improvement in RMS error.

**Table 4. Improvement measures for trained stimulus in the echoic environment. Mean values across subjects in the visually-trained group are presented along with the standard deviation reflecting the inter-subject variability.**



| | Precision | Accuracy | RMS |
|---|---|---|---|
| *EI* | 15.5° ± 8.6° | 14.9° ± 12.3° | 21.4° ± 14.4° |
| *NEI* | 60% ± 26% | 63% ± 42% | 58% ± 34% |

The pre-training and post-training errors of all types can be seen in Table 5 and Table 6 for the trained stimulus (sine tone) and in Table 7 and Table 8 for the untrained stimulus (white noise burst).

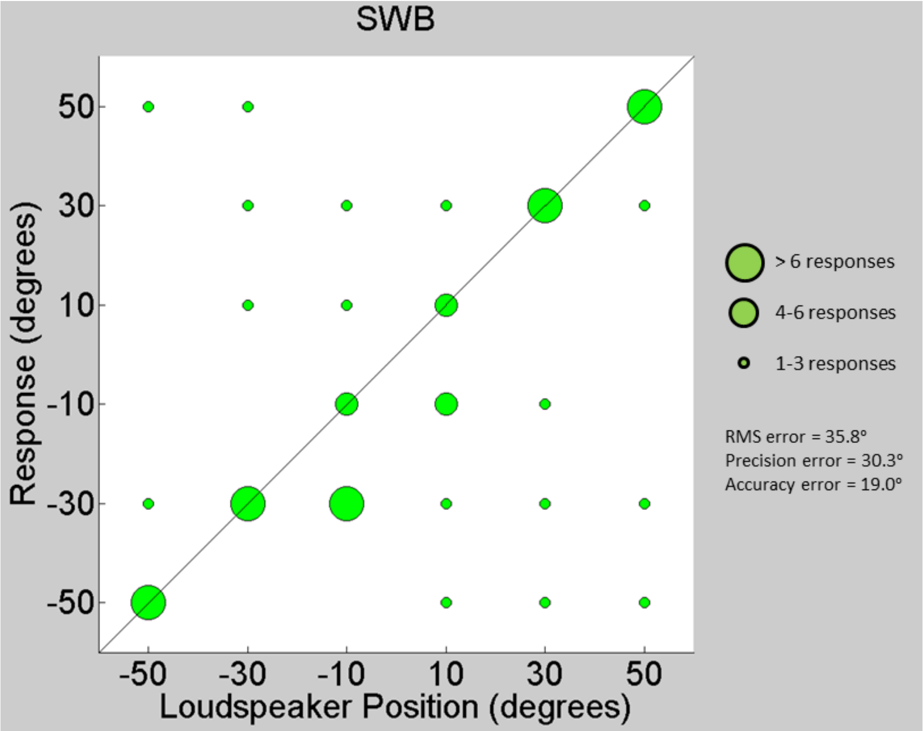


**Figure 5. Pre-training localization data for subject SWB with the sine tone stimulus. A perfect localization performance corresponds to having all data points on the main diagonal. Bigger symbols denote higher frequency of response to that position.**

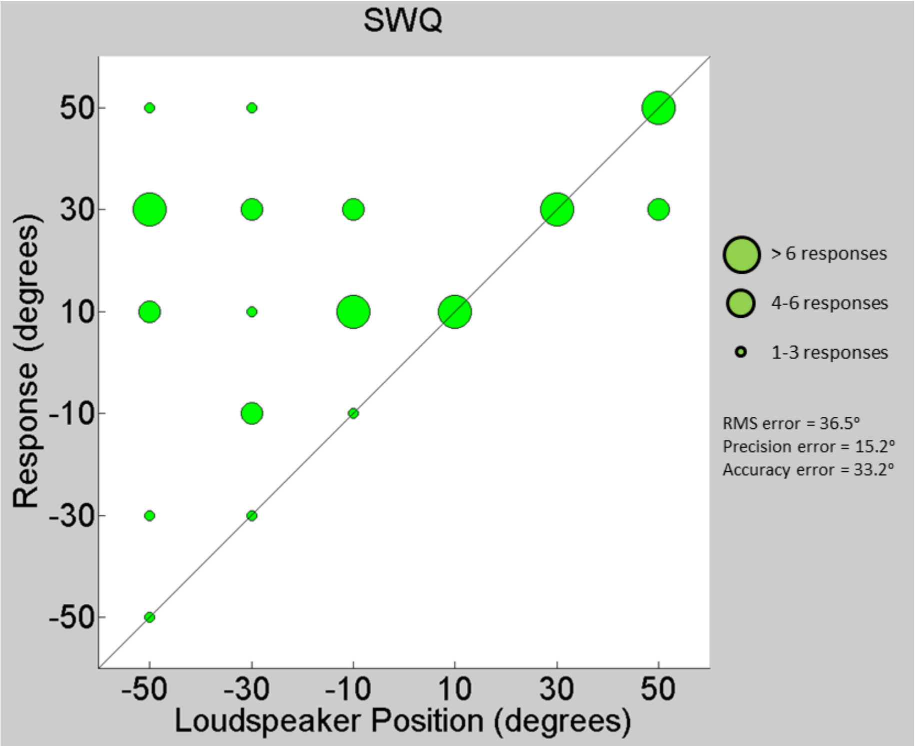
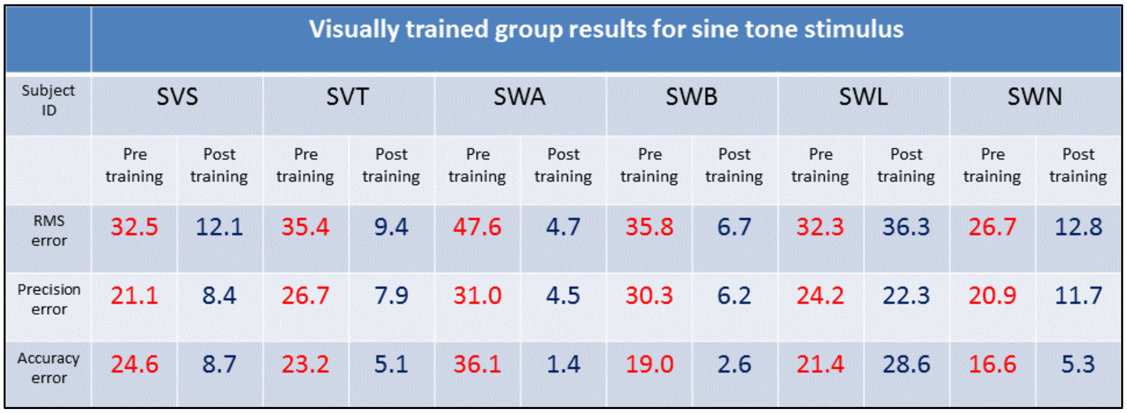


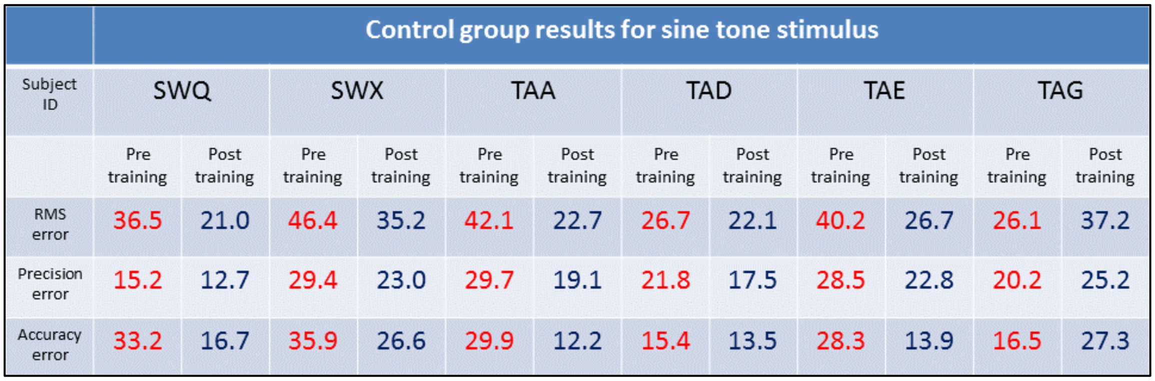
**Figure 6. Pre-training localization data for subject SWB with the sine tone stimulus. Bigger symbols denote higher frequency of response to that position.**

Table 5. Pre- and post-training errors (shown in units of degrees) for six visually-trained listeners tested on the sine tone stimulus.



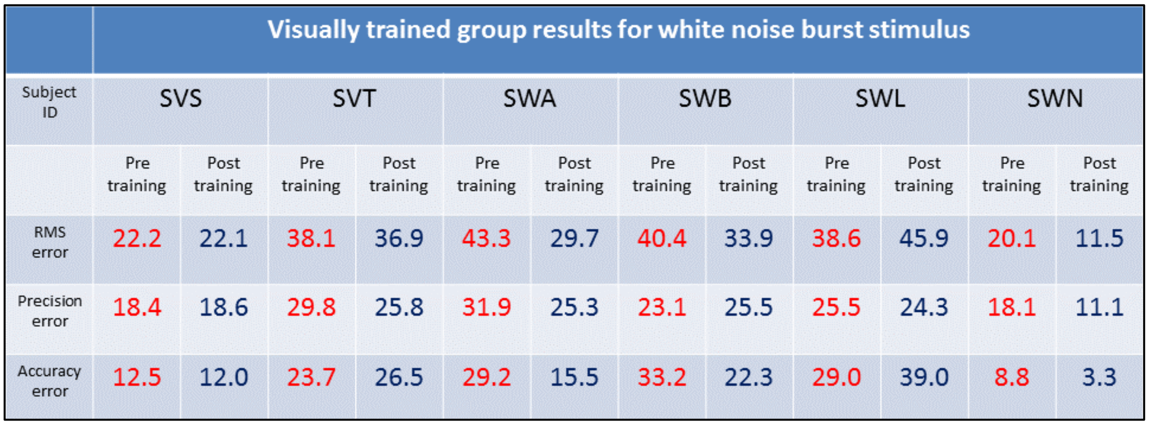
| | Visually trained group results for sine tone stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | SVS | | SVT | | SWA | | SWB | | SWL | | SWN | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 32.5 | 12.1 | 35.4 | 9.4 | 47.6 | 4.7 | 35.8 | 6.7 | 32.3 | 36.3 | 26.7 | 12.8 |
| Precision error | 21.1 | 8.4 | 26.7 | 7.9 | 31.0 | 4.5 | 30.3 | 6.2 | 24.2 | 22.3 | 20.9 | 11.7 |
| Accuracy error | 24.6 | 8.7 | 23.2 | 5.1 | 36.1 | 1.4 | 19.0 | 2.6 | 21.4 | 28.6 | 16.6 | 5.3 |

Table 6. Pre- and post-training errors (shown in units of degrees) for six control listeners tested on the sine tone stimulus.



| | Control group results for sine tone stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | SWQ | | SWX | | TAA | | TAD | | TAE | | TAG | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 36.5 | 21.0 | 46.4 | 35.2 | 42.1 | 22.7 | 26.7 | 22.1 | 40.2 | 26.7 | 26.1 | 37.2 |
| Precision error | 15.2 | 12.7 | 29.4 | 23.0 | 29.7 | 19.1 | 21.8 | 17.5 | 28.5 | 22.8 | 20.2 | 25.2 |
| Accuracy error | 33.2 | 16.7 | 35.9 | 26.6 | 29.9 | 12.2 | 15.4 | 13.5 | 28.3 | 13.9 | 16.5 | 27.3 |

Table 7. Pre- and post-training errors (shown in units of degrees) for six visually-trained listeners tested on the white noise burst stimulus.



| | Visually trained group results for white noise burst stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | SVS | | SVT | | SWA | | SWB | | SWL | | SWN | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 22.2 | 22.1 | 38.1 | 36.9 | 43.3 | 29.7 | 40.4 | 33.9 | 38.6 | 45.9 | 20.1 | 11.5 |
| Precision error | 18.4 | 18.6 | 29.8 | 25.8 | 31.9 | 25.3 | 23.1 | 25.5 | 25.5 | 24.3 | 18.1 | 11.1 |
| Accuracy error | 12.5 | 12.0 | 23.7 | 26.5 | 29.2 | 15.5 | 33.2 | 22.3 | 29.0 | 39.0 | 8.8 | 3.3 |

Table 8. Pre- and post-training errors (shown in units of degrees) for six control listeners tested on the white noise burst stimulus.

| | Control group results for white noise burst stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | SWQ | | SWX | | TAA | | TAD | | TAE | | TAG | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 35.3 | 42.7 | 47.0 | 40.1 | 38.5 | 41.7 | 38.2 | 29.1 | 41.1 | 31.3 | 33.1 | 40.6 |
| Precision error | 24.4 | 35.8 | 28.6 | 23.3 | 25.2 | 26.6 | 22.9 | 22.2 | 27.9 | 25.0 | 24.9 | 27.9 |
| Accuracy error | 25.5 | 23.3 | 37.2 | 32.6 | 29.1 | 32.3 | 30.6 | 18.9 | 30.1 | 18.8 | 21.9 | 29.5 |

## D.2. Cause of improvement

The performance of the visually-trained group was compared to that of the control group in order to assess the influence of the AVIT paradigm on the observed improvements in localization ability. Kruskal-Wallis tests were used to compare the error improvement of the two groups. The statistics in Table 9 suggest that the AVIT paradigm had a significant effect in improving both precision and accuracy for the trained stimulus. Some subjects exhibited larger accuracy improvements, such as SWA (Figure 7), while others exhibited larger precision improvements, such as SWB (Figure 8).

Table 9. Statistics for a Kruskal-Wallis test between the improvements of the visually-trained group and the control group. Values are presented across different improvement measures for the sine tone stimulus. Values highlighted in yellow are considered statistically significant in psychophysics literature (p<0.05)

| Comparison between visually trained group and control group | Precision Improvement | Normalized Precision Improvement | Accuracy Improvement | Normalized Accuracy Improvement | RMS Improvement | Normalized RMS Improvement |
|---|---|---|---|---|---|---|
| | H(1,10)= 3.69<br>p = 0.0547 | H(1,10) = 4.33<br>p = 0.0374 | H(1,10) = 1.09<br>p = 0.2971 | H(1,10) = 4.33<br>p = 0.0374 | H(1,10) = 3.1<br>p = 0.0782 | H(1,10) = 4.33<br>p = 0.0374 |

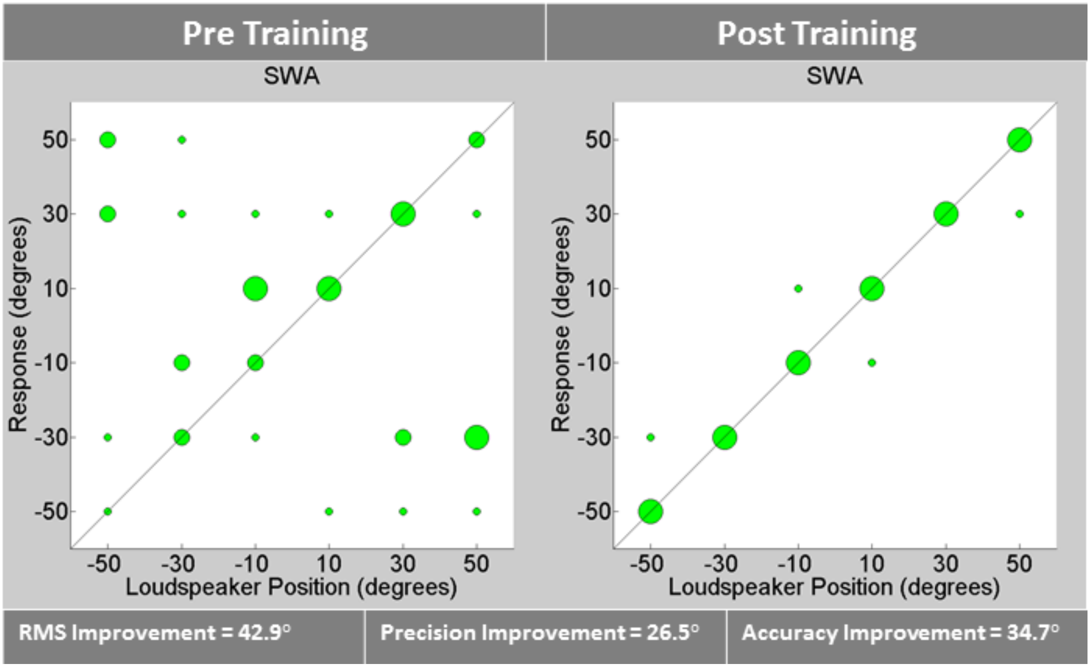


**Figure 7. Pre-training and post-training localization data for subject SWA tested on the sine tone stimulus.**

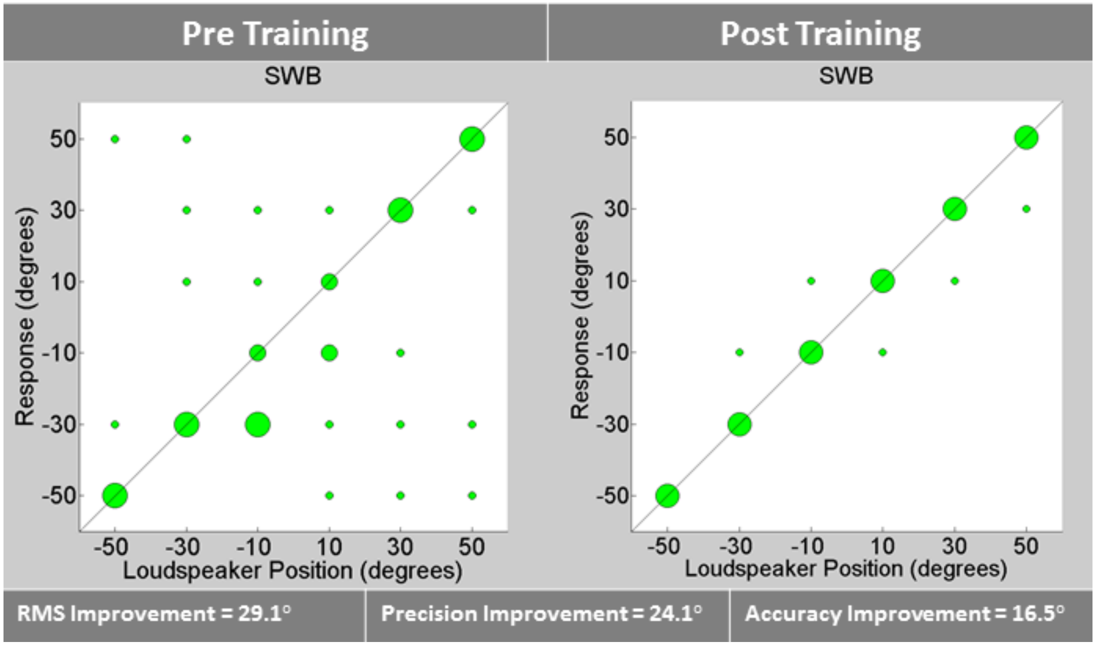


**Figure 8. Pre-training and post-training localization data for subject SWB tested on the sine tone stimulus**

### D.3. Generalization

Kruskal-Wallis tests comparing the pre-training and post-training errors (see appendix B, Table 21) indicated no significant improvement for the white noise burst stimulus, suggesting that learning due to AVIT did not generalized to an untrained stimulus. This points to the specialized manner of the precedence effect.  The PE has been shown to be sensitive to environmental changes as delicate as switching the location of the leading and lagging sound [40,60,61].

Another point of interest was whether learning would generalize to untrained target positions. Generalization could be tested only for those subjects who exhibited improvement. Kruskal-Wallis tests for the fine-grained testing stage, showed no significant difference between the subjects' performance on spatial locations for which they received training and for which they did not (see appendix B, Table 23). It is not possible to conclude that learning generalized to untrained target locations due to the lack of pre-training fine-grained testing data. However, the AVIT method can be claimed to have generated a continuous change of the subjects' auditory spatial map rather than a mere memorization of spatial locations.

AVIT was shown to be an effective method to improve localization ability in the echoic environment described in experiment 1. Listeners were able to suppress lagging soundwaves more effectively after being trained with a reinforcing visual stimulus, which coincided spatially and temporally with the leading soundwave. The visual stimulus allowed the listeners to weight the auditory cues associated with the leading soundwave more heavily than those associated with the lagging soundwaves. The type of auditory cues associated with each soundwave were the same (ILDs, ITDs and spectral cues). In the experiment described below, the efficacy of AVIT is examined in a different environment. An environment where a single sound, containing both salient and degraded auditory cues, is presented to the listeners. In this environment AVIT may allow the listeners to weight the salient auditory cues (ILDs) more heavily than the degraded auditory cues (ITDs and spectral cues).

# III. EXPERIMENT 2: LOCALIZATION OF VOCODED STIMULI

Experiment 2 aims to test the hypothesis that AVIT can allow listeners to weight salient localization cues more heavily than degraded localization cues, thereby improving the listeners' localization ability. The testing environment in this experiment aims to simulate the effect a CI has on a natural stimulus with salient ILD, ITD and spectral cues. Therefore, the stimulus used in this experiment contains ILD cues that are more reliable than ITD or spectral cues. Confirming this experiment's hypothesis would provide further supportive evidence of the potential benefit AVIT has for CI user rehabilitation.

## A. Listeners and equipment

Eleven NH adult subjects (mean age 26, range 19-47) participated in the study. None participated in experiment 1. They were divided into a visually-trained group (six subjects) and a control group (five subjects). All subjects reported no auditory or neurological disease and had normal or corrected to normal vision. The listeners had pure tone thresholds within 20 dB of audiometrically normal hearing at octave interval frequencies between 250 and 8000 Hz, and the thresholds between the two ears differed by less than 15 dB at any tested frequency.

The equipment used in experiment 2 was identical to that used in experiment 1. HRTFs were individually measured by using a blocked-meatus microphone pair (HeadZap binaural microphones, AuSIM Inc.) placed in the entrance of each ear canal. Auditory stimuli were delivered to the listener through ER-2 headphones.

## B. Stimuli

Stimuli included vocoded and natural versions of the Consonant-Vowel-Consonant (CVC) words *Merge* and *Beam*, both at 60 dB. The natural versions of the stimuli contained ILD cues, ITD cues both in the envelope and fine structure, and spectral cues. The natural stimuli were convolved with the subject's HRTF in order to create a virtual acoustic environment [62] and then either vocoded or left in its natural form. The vocoding process followed the general scheme described by Shannon et al. (1995), with 8 analysis and output filters between 150 Hz and 8000 Hz, according to the Greenwood function that correlates the position of hair cells to the frequency of their corresponding auditory neurons [63]. Envelope extraction was performed via half-wave rectification and low pass filtering at 50 Hz.

The vocoding process degraded the envelope and fine structure ITD cues, as well as any spectral cues. Carriers were interaurally decorrelated in order to avoid an artificial ITD cue across different frequencies that might not be available to a CI user. This was done by performing a Grahm-Schmidt orthogonalization process on the left and right white noise carriers. Three different versions of the stimuli were constructed, each from a different random white noise distribution, in order to assure the lack of an artificial ITD cue. ILDs were not degraded in any significant way as can be seen from Figure 9 and Figure 10. Spectrograms of both natural and vocoded stimuli can be seen in Figure 11 where a temporal envelope cue across frequency bands is clearly seen in the natural stimulus but is absent in the vocoded stimulus.

Virtual environment natural stimuli has been shown to produce similar localization performance to that of NH listeners localizing sounds arriving from physical loudspeakers in a sound booth, known as *free field* conditions [15,64]. Vocoding virtual environment stimuli has been shown to produce similar performance to that of CI users localizing natural stimuli in free field conditions [64].

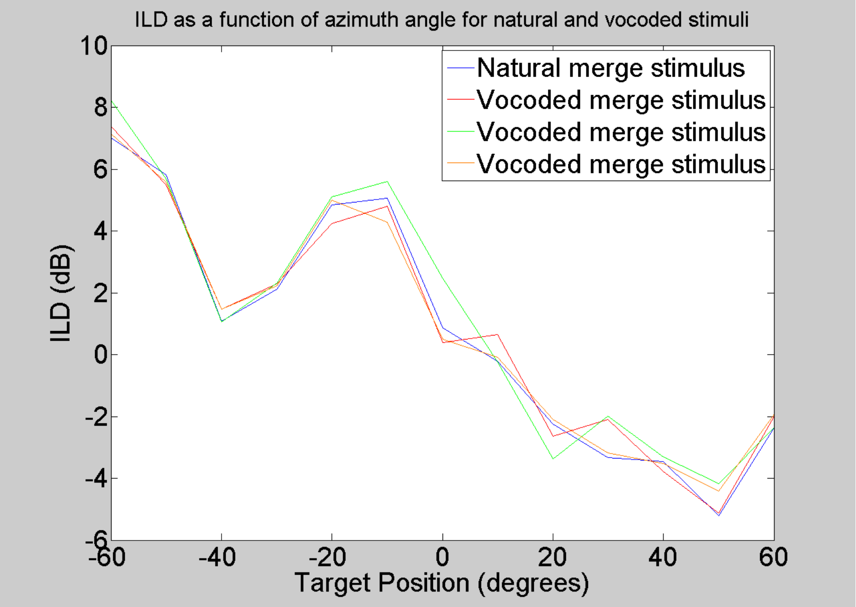


**Figure 9. Three versions of the vocoded merge stimulus, each constructed with a different random white noise carrier. The ILD functions of these 3 versions are plotted against the natural version of the stimulus, showing no significant difference.**

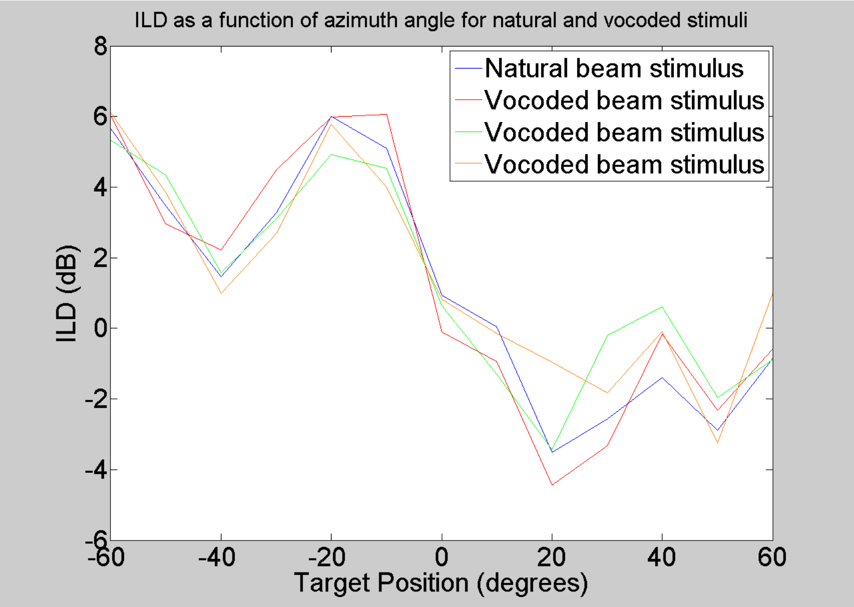


**Figure 10. Three versions of the vocoded beam stimulus, each constructed with a different random white noise carrier. The ILD functions of these 3 versions are plotted against the natural version of the stimulus, showing significant difference.**

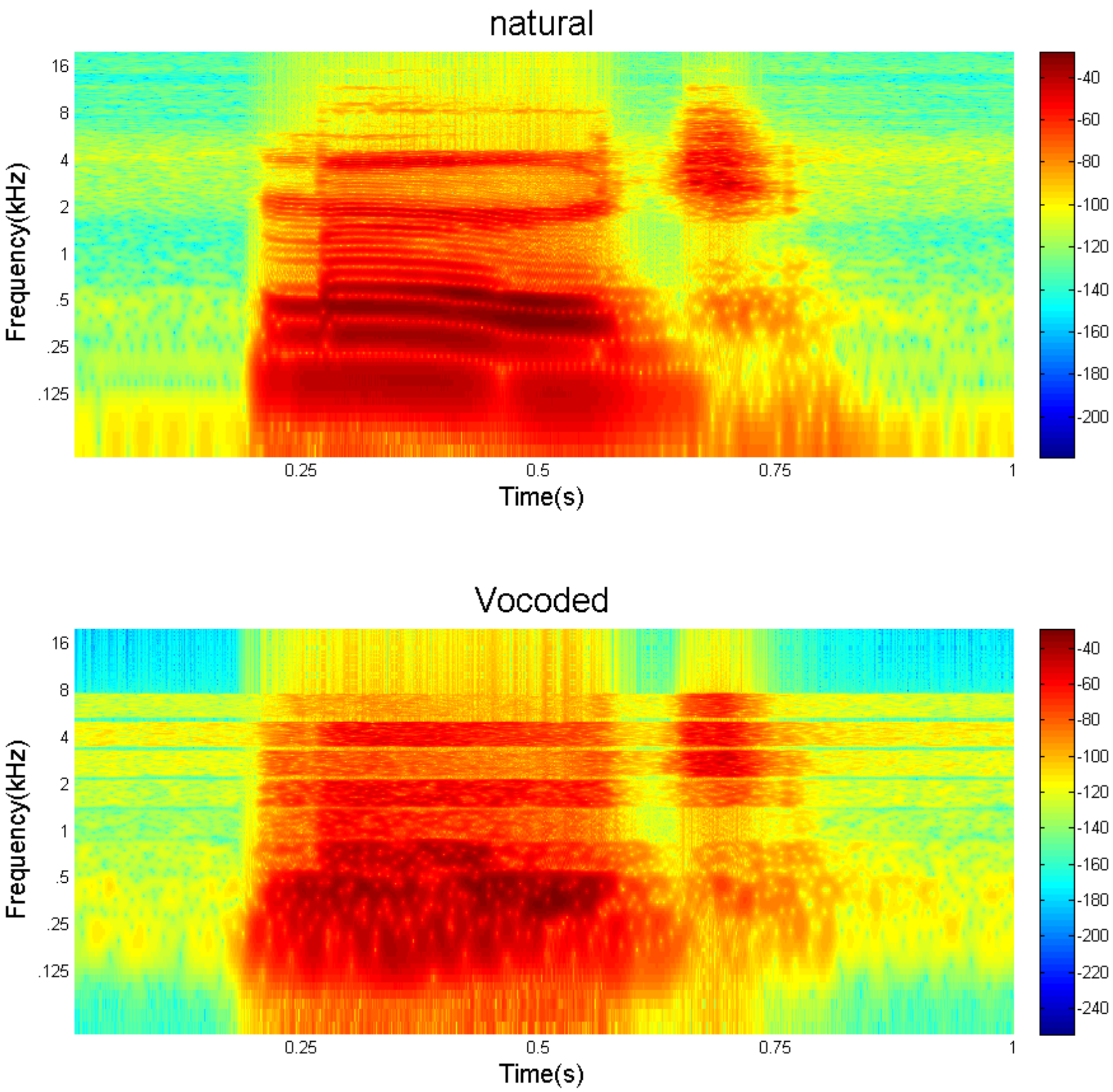
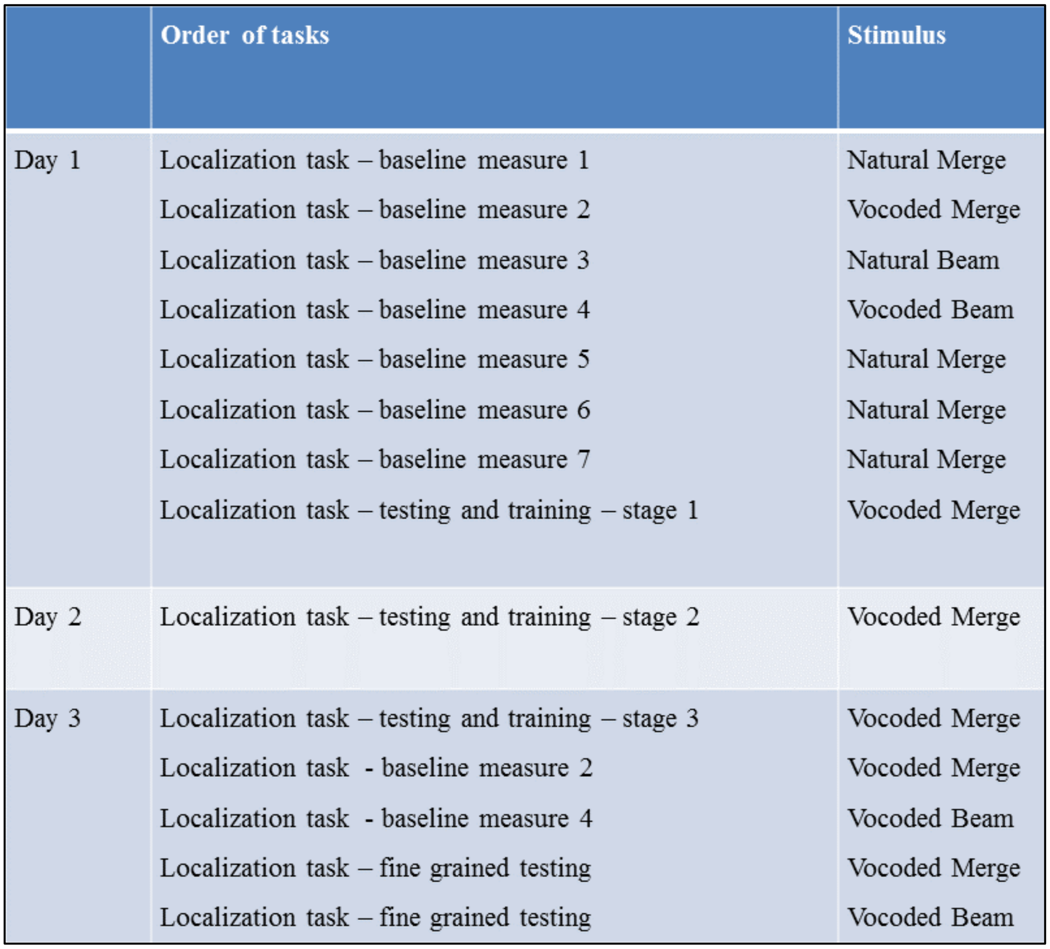


Figure 11. Spectrograms of the natural (unprocessed) and vocoded word merge. The color scheme represents the amplitude of the sound in Decibels. An envelope modulation across time can be seen in many frequency bands on the top spectrogram (especially between 250 Hz and 500 Hz) but not on the bottom spectrogram.

## C. Procedure

Subject participation in the study took place on 3 separate days, all within a 7-day time span. Each day participants were tested and trained on a localization task. The number of loudspeakers in the task increased from day to day in order to promote a gradual training scenario. Subjects in the visually-trained group received a reinforcing visual stimulus during training. This visual stimulus was absent during the training stage of the control group. The subjects' localization ability was measured pre- and post-training in order to assess improvement. The order of testing in the 3 different study days is summarized in Table 10. The hypothesis was that visually-trained subjects would show statistically significant improvement in localization ability compared to subjects in the control group. This is hypothesized to be a result of listeners weighting ILD cues more heavily post-training.

Table 10. Order of testing for experiment 2.

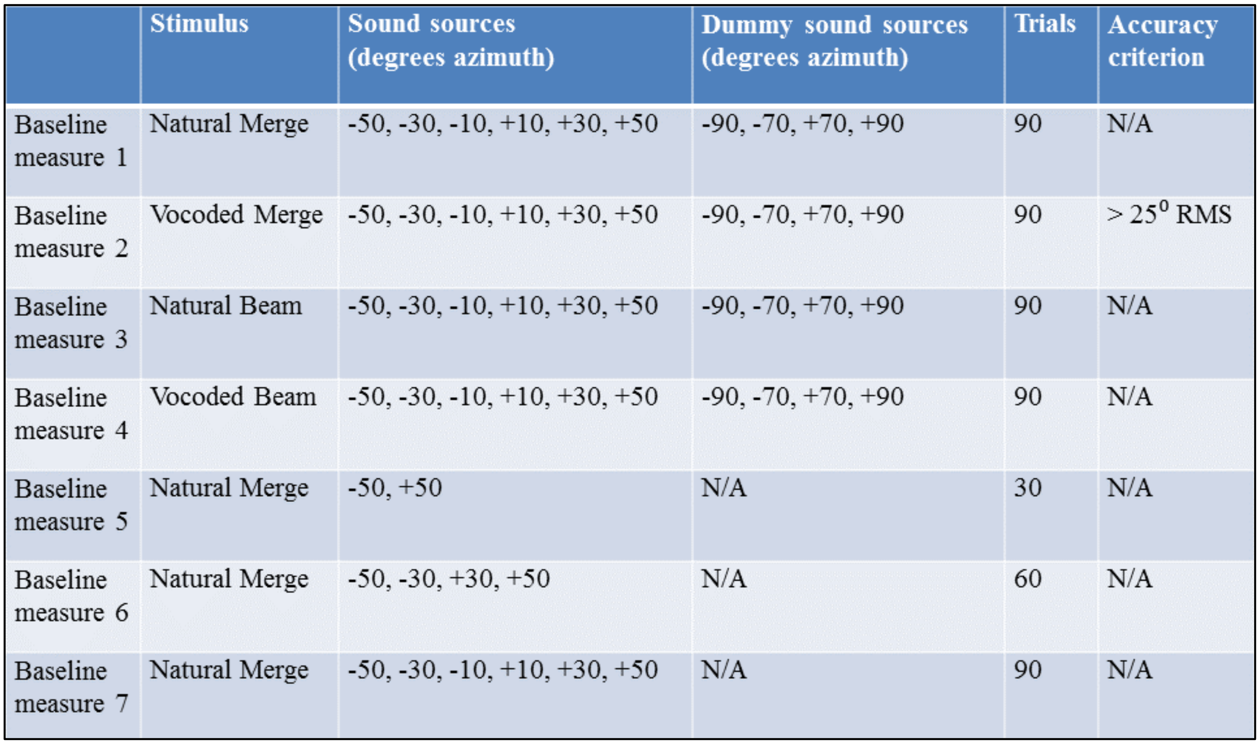


| | Order of tasks | Stimulus |
|---|---|---|
| Day 1 | Localization task – baseline measure 1 | Natural Merge |
| | Localization task – baseline measure 2 | Vocoded Merge |
| | Localization task – baseline measure 3 | Natural Beam |
| | Localization task – baseline measure 4 | Vocoded Beam |
| | Localization task – baseline measure 5 | Natural Merge |
| | Localization task – baseline measure 6 | Natural Merge |
| | Localization task – baseline measure 7 | Natural Merge |
| | Localization task – testing and training – stage 1 | Vocoded Merge |
| Day 2 | Localization task – testing and training – stage 2 | Vocoded Merge |
| Day 3 | Localization task – testing and training – stage 3 | Vocoded Merge |
| | Localization task - baseline measure 2 | Vocoded Merge |
| | Localization task - baseline measure 4 | Vocoded Beam |
| | Localization task – fine grained testing | Vocoded Merge |
| | Localization task – fine grained testing | Vocoded Beam |

**Localization task**

The localization portion consisted of 7 baseline measurements, as well as testing and training that gradually increased in difficulty and duration from one day to the other. All 7 baseline measurements corresponded to a 6-source configuration (-50°,-30°,-10°,+10°,+30°,+50° azimuth) or a subset of it. Subjects were instructed to indicate the position of the sound during baseline measurements and testing. Details of the baseline measures can be found in Table 11.

Table 11. Baseline measure setup for experiment 2.



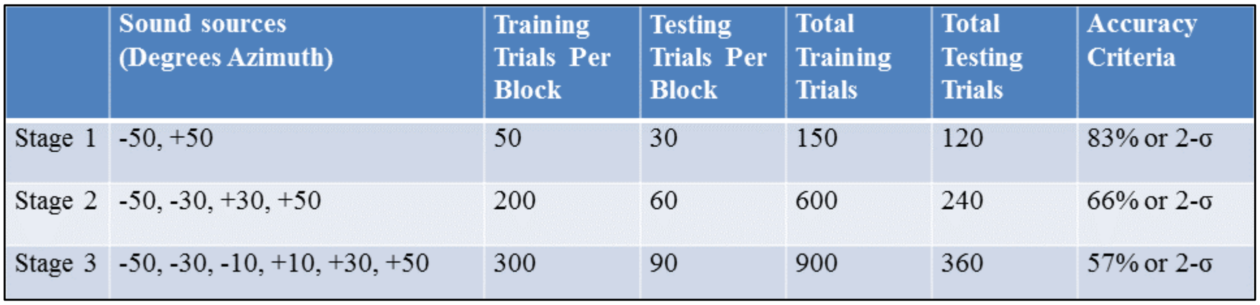
| | Stimulus | Sound sources (degrees azimuth) | Dummy sound sources (degrees azimuth) | Trials | Accuracy criterion |
|---|---|---|---|---|---|
| Baseline measure 1 | Natural Merge | -50, -30, -10, +10, +30, +50 | -90, -70, +70, +90 | 90 | N/A |
| Baseline measure 2 | Vocoded Merge | -50, -30, -10, +10, +30, +50 | -90, -70, +70, +90 | 90 | > $25^0$ RMS |
| Baseline measure 3 | Natural Beam | -50, -30, -10, +10, +30, +50 | -90, -70, +70, +90 | 90 | N/A |
| Baseline measure 4 | Vocoded Beam | -50, -30, -10, +10, +30, +50 | -90, -70, +70, +90 | 90 | N/A |
| Baseline measure 5 | Natural Merge | -50, +50 | N/A | 30 | N/A |
| Baseline measure 6 | Natural Merge | -50, -30, +30, +50 | N/A | 60 | N/A |
| Baseline measure 7 | Natural Merge | -50, -30, -10, +10, +30, +50 | N/A | 90 | N/A |

Baseline measures 2 and 4 allow calculating localization improvement by comparing a subject's performance on them pre- and post-training. "Dummy" sound sources appear as response options for the listener while no sound actually originate from their direction. The purpose of the "dummy" sound sources is to minimize edge effects (caused by the limited allowable response range) and increase the probability that a listener would perform poorly prior to training. A localization task that spans a larger range of targets, has a larger RMS error at *chance performance* and is thus considered more difficult.

A subject's performance on the first and third baseline measures allows estimating the highest level of performance that can be induced post-training, as the natural stimulus contains more localization cues than the vocoded stimulus. Baseline measures 5, 6 and 7 correspond to the three training stages and allow the setting of one performance criteria for each stage: An RMS error within two standard deviations of the mean RMS error in the corresponding baseline measure. This criterion is denoted 2-σ.

Testing and training were performed gradually with one stage for each day (Table 10). Subjects in the visually-trained group were trained with a one second LED flash originating from the location of and simultaneously with the sound source. The subjects were instructed to indicate the position of the light while still trying to locate the sound. Three training blocks were alternated with four testing blocks (in which no visual stimulus was present) in order to assess the subject's improvement from one training block to the next. Data detailing the progress of each listener throughout the study procedure can be found in appendix C. Only subjects in the visually-trained group were required to perform to a predetermined level in order to advance from one day to the next. Details of the training stages as well as the performance criteria associated with them are shown in Table 12.

Table 12. Setup of training stages in experiment 2.



| | Sound sources (Degrees Azimuth) | Training Trials Per Block | Testing Trials Per Block | Total Training Trials | Total Testing Trials | Accuracy Criteria |
|---|---|---|---|---|---|---|
| Stage 1 | -50, +50 | 50 | 30 | 150 | 120 | 83% or 2-σ |
| Stage 2 | -50, -30, +30, +50 | 200 | 60 | 600 | 240 | 66% or 2-σ |
| Stage 3 | -50, -30, -10, +10, +30, +50 | 300 | 90 | 900 | 360 | 57% or 2-σ |

The last localization task in the experiment, denoted *fine grained testing* consisted of 180 trials in a 12 sound source configuration (Targets at -60°,-50°,-40°,-30°,-20°,-10°,+10°,+20°,+30°,+40°,+50°,+60° azimuth plus "dummy" targets at -90°, -80°, -70°, 0°, +70°, +80°, +90°). It was performed first for the vocoded Merge stimulus and then for the vocoded Beam stimulus. During the *fine grained testing* task subjects localize sounds originating from locations for which they received training and for which they did not. Comparing the subjects' performance on these two types of source locations will shed light on the type of effect the training had on their auditory maps of space. Specifically, the *fine grained testing* task will allow determining whether subjects merely memorized the relationship between the stimulus and the trained locations, or if their auditory maps of space were altered in a continuous way.

The gradual training protocol differs between the first and second experiment. In the former, training starts with loudspeakers at -10° and +10° azimuth and expands sideways, whereas in the latter, training starts with loudspeaker at -50° and +50° azimuth and expands towards the middle. The latter protocol has a less rigorous starting stage as the large separation between the loudspeakers makes for easier discrimination. This enabled more subjects to meet performance criteria, and thus transition onto later training stages. The number of training trials in the first stage in Table 12 is relatively small because of its low level of difficulty (subjects reached performance criteria quickly).

## D. Results

The same data analysis methods presented in experiment 1, were used for the topics of dominant error types, cause of improvement, and generalization to untrained condition for this experiment.

### D.1. Dominant error types

The pre-training precision errors of all subjects (visually-trained and control groups) were compared with the pre-training accuracy errors of all subjects via a Kruskal-Wallis test. The reported statistics in Table 13 shows that accuracy errors were the dominant error type for both the trained ('merge') and untrained ('beam') stimuli. This can be clearly seen in subject TCE that exhibited good precision but poor accuracy (Figure 12). As expected the visually-trained group exhibited higher improvement measures on accuracy than precision for the trained stimulus. All the error improvement measures can be seen in Table 14. The pre-training and post-training errors of all types can be seen in Table 15 and Table 16 for the trained stimulus and in Table 17 and Table 18 for the untrained stimulus.

Table 13. Statistics for a Kruskal-Wallis test between the pre-training accuracy error of all subjects and the pre-training precision error of all subjects. All p-values are below 0.05, the statistically significant benchmark in the psychophysics literature. Pre-training errors presented include mean values across all twelve subjects as well as standard deviations.

| Merge | | Beam | |
|---|---|---|---|
| Precision | Accuracy | Precision | Accuracy |
| 21.7° ± 9.0° | 31.3° ± 11.1° | 19.4° ± 4.9° | 34.5 ± 11.1° |
| H(1,20) = 5.74, p = 0.0165 | | H(1,20) = 11.88, p = 0.0006 | |

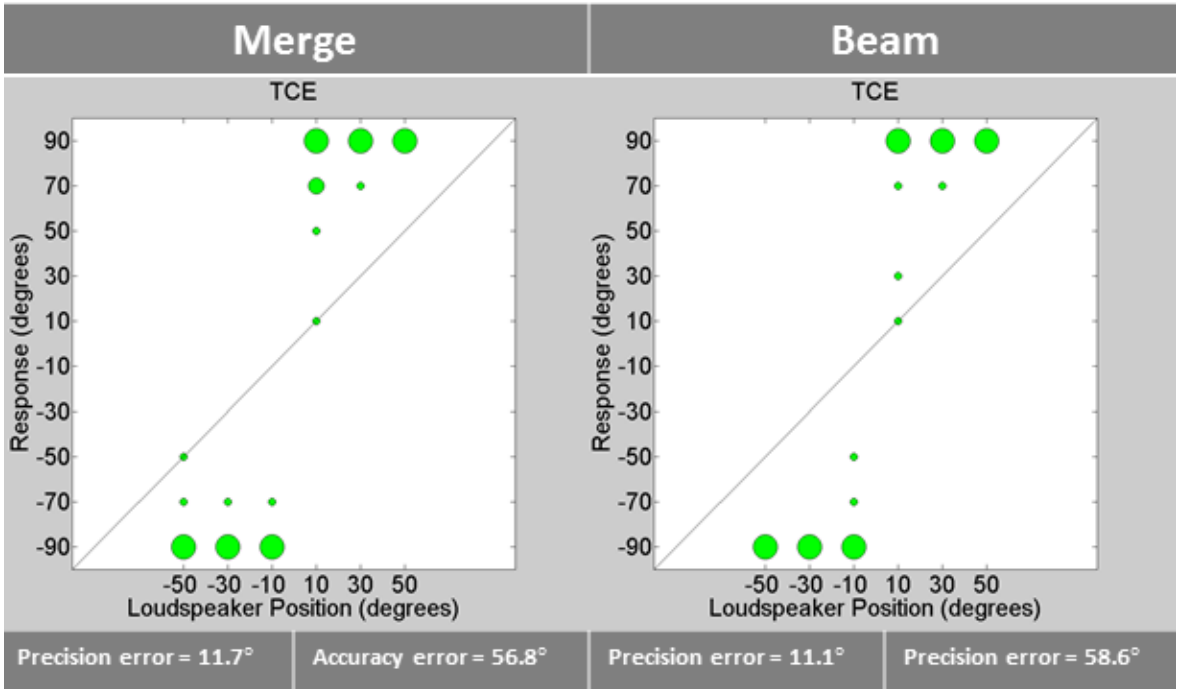
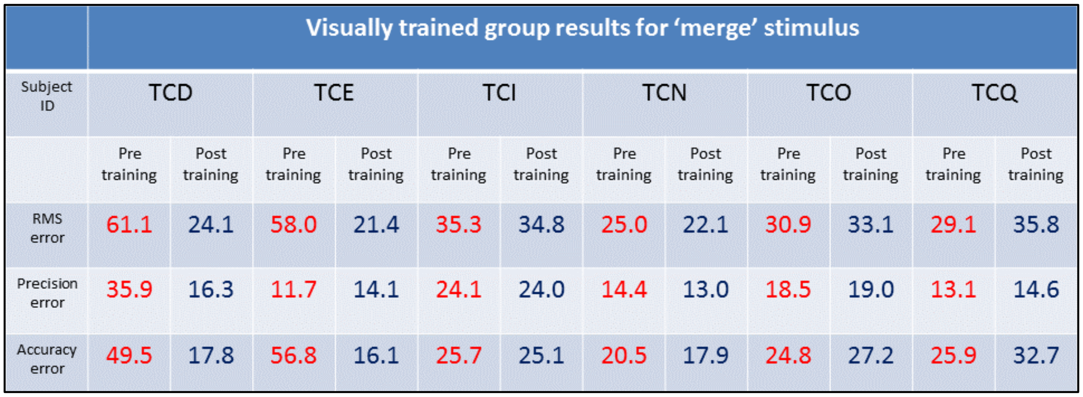


Figure 12. Pre-training localization data for subject TCE.

Table 14. Improvement measures for trained stimulus in the CI simulated environment. Mean values across subjects in the visually-trained group are presented along with the standard deviation reflecting the inter-subject variability.

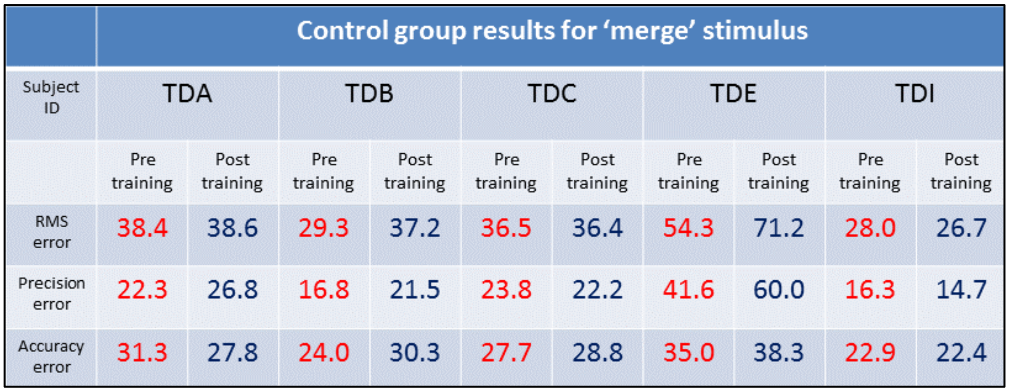
| | Precision | Accuracy | RMS |
|---|---|---|---|
| *EI* | 2.8° ± 7.6° | 11.1° ± 18.2° | 11.3° ± 18.3° |
| *NEI* | 14% ± 24% | 33% ± 36% | 18% ± 33% |

**Table 15. Pre- and post-training errors (shown in units of degrees) for six visually-trained listeners tested on the merge stimulus.**



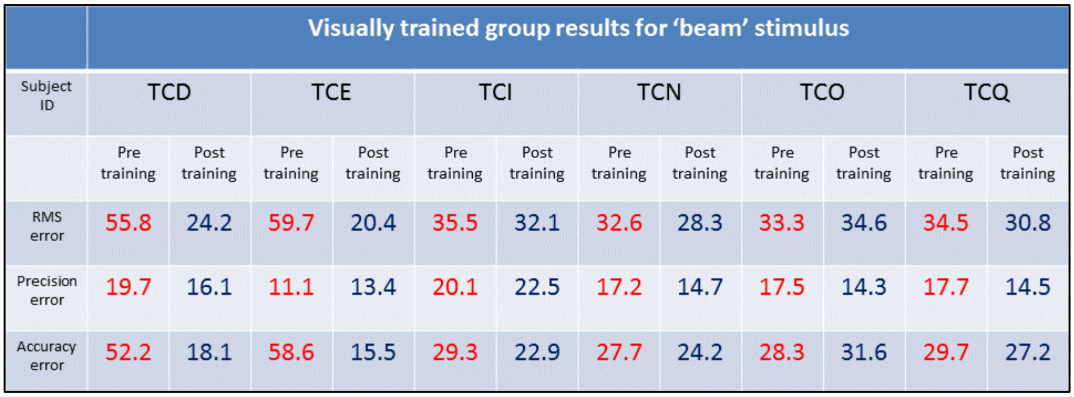
| | Visually trained group results for 'merge' stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | TCD | | TCE | | TCI | | TCN | | TCO | | TCQ | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 61.1 | 24.1 | 58.0 | 21.4 | 35.3 | 34.8 | 25.0 | 22.1 | 30.9 | 33.1 | 29.1 | 35.8 |
| Precision error | 35.9 | 16.3 | 11.7 | 14.1 | 24.1 | 24.0 | 14.4 | 13.0 | 18.5 | 19.0 | 13.1 | 14.6 |
| Accuracy error | 49.5 | 17.8 | 56.8 | 16.1 | 25.7 | 25.1 | 20.5 | 17.9 | 24.8 | 27.2 | 25.9 | 32.7 |

**Table 16. Pre- and post-training errors (shown in units of degrees) for five control listeners tested on the merge stimulus.**



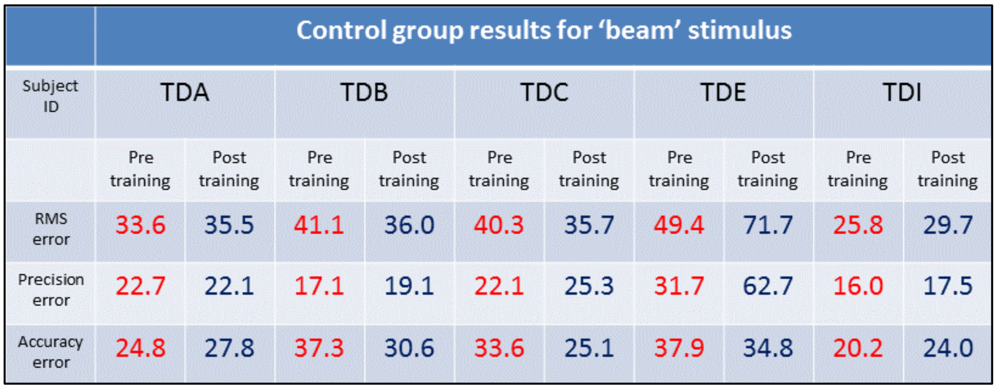
| | Control group results for 'merge' stimulus | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | TDA | | TDB | | TDC | | TDE | | TDI | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 38.4 | 38.6 | 29.3 | 37.2 | 36.5 | 36.4 | 54.3 | 71.2 | 28.0 | 26.7 |
| Precision error | 22.3 | 26.8 | 16.8 | 21.5 | 23.8 | 22.2 | 41.6 | 60.0 | 16.3 | 14.7 |
| Accuracy error | 31.3 | 27.8 | 24.0 | 30.3 | 27.7 | 28.8 | 35.0 | 38.3 | 22.9 | 22.4 |

**Table 17. Pre- and post-training errors (shown in units of degrees) for six visually-trained listeners tested on the beam stimulus.**



| | Visually trained group results for 'beam' stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | TCD | | TCE | | TCI | | TCN | | TCO | | TCQ | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 55.8 | 24.2 | 59.7 | 20.4 | 35.5 | 32.1 | 32.6 | 28.3 | 33.3 | 34.6 | 34.5 | 30.8 |
| Precision error | 19.7 | 16.1 | 11.1 | 13.4 | 20.1 | 22.5 | 17.2 | 14.7 | 17.5 | 14.3 | 17.7 | 14.5 |
| Accuracy error | 52.2 | 18.1 | 58.6 | 15.5 | 29.3 | 22.9 | 27.7 | 24.2 | 28.3 | 31.6 | 29.7 | 27.2 |

Table 18. Pre- and post-training errors (shown in units of degrees) for five control listeners tested on the beam stimulus.

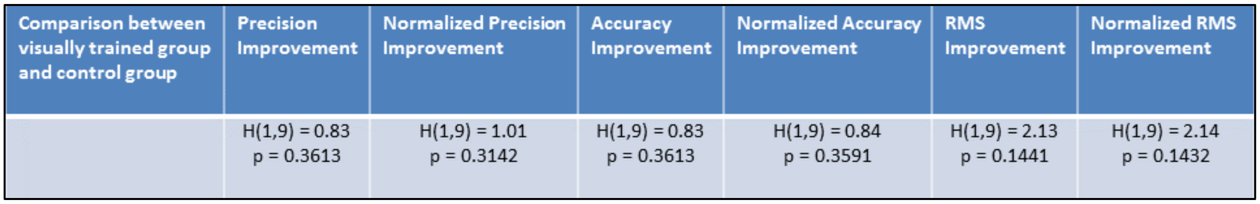


| | Control group results for 'beam' stimulus | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Subject ID | TDA | | TDB | | TDC | | TDE | | TDI | |
| | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training | Pre training | Post training |
| RMS error | 33.6 | 35.5 | 41.1 | 36.0 | 40.3 | 35.7 | 49.4 | 71.7 | 25.8 | 29.7 |
| Precision error | 22.7 | 22.1 | 17.1 | 19.1 | 22.1 | 25.3 | 31.7 | 62.7 | 16.0 | 17.5 |
| Accuracy error | 24.8 | 27.8 | 37.3 | 30.6 | 33.6 | 25.1 | 37.9 | 34.8 | 20.2 | 24.0 |

## D.2. Cause of improvement

The performance of the visually-trained group was compared to that of the control group in order to assess the influence of the AVIT paradigm on the observed improvements in localization ability. Statistical testing with Kruskal-Wallis tests was used to compare the error improvement of the two groups. The statistics in Table 19 suggest that the AVIT paradigm had no significant effect in improving localization for the trained stimulus. However, some visually-trained subjects like TCD (Figure 13) and TCE (Figure 14) exhibited large improvements, especially in accuracy.

Table 19. Statistics for a Kruskal-Wallis test comparing the improvements of the visually-trained group and the control group. Values are presented across different improvement measures for the merge stimulus



| Comparison between visually trained group and control group | Precision Improvement | Normalized Precision Improvement | Accuracy Improvement | Normalized Accuracy Improvement | RMS Improvement | Normalized RMS Improvement |
|---|---|---|---|---|---|---|
| | H(1,9) = 0.83<br>p = 0.3613 | H(1,9) = 1.01<br>p = 0.3142 | H(1,9) = 0.83<br>p = 0.3613 | H(1,9) = 0.84<br>p = 0.3591 | H(1,9) = 2.13<br>p = 0.1441 | H(1,9) = 2.14<br>p = 0.1432 |

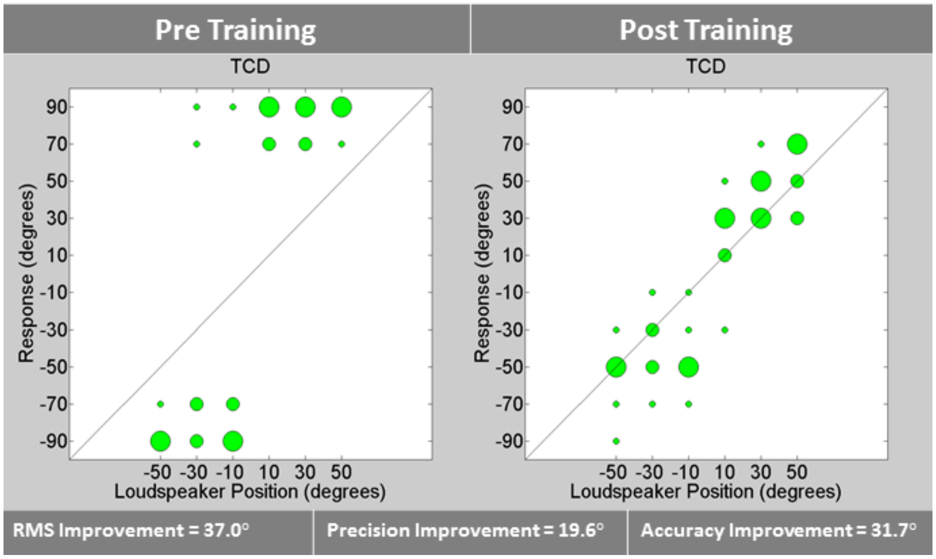


Figure 13. Pre- training and post-training localization data for subject TCD tested on the merge stimulus.

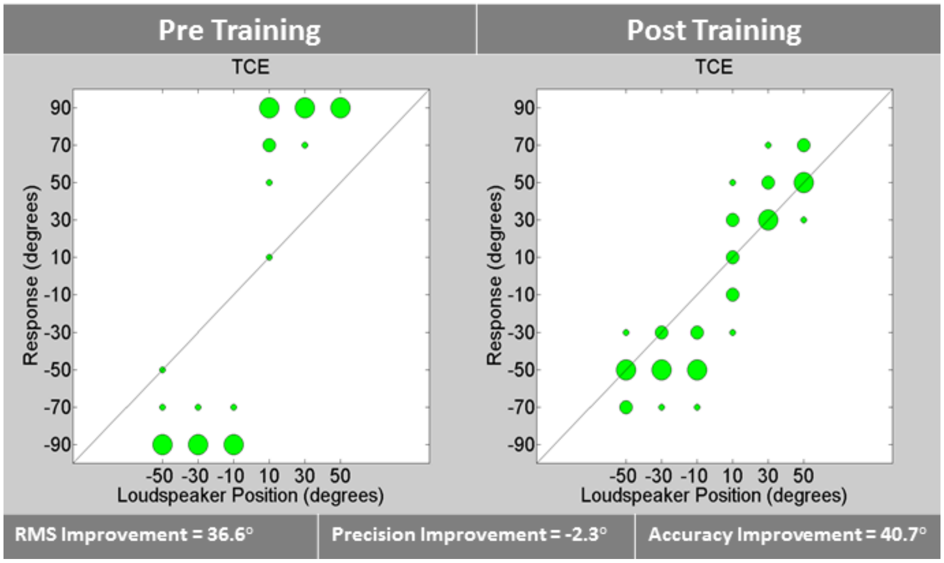
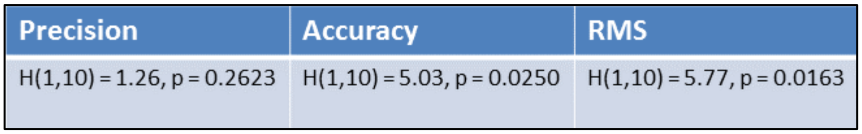


Figure 14. Pre- training and post-training localization data for subject TCE tested on the merge stimulus.

### D.3. Generalization

Kruskal-Wallis tests comparing the pre-training and post-training errors of visually-trained subjects showed significant improvement in accuracy but not in precision for the untrained stimulus (Table 20). However, no statistically significant improvement was found when the visually-trained group performance was compared to that of the control group (see appendix B, Table 22), suggesting that learning due to AVIT did not generalized to an untrained stimulus. Nonetheless, some visually-trained subjects like TCD (Figure 15) and TCE (Figure 16) exhibited large improvements for the untrained stimulus, especially in accuracy.

Table 20. Kruskal-Wallis statistics comparing pre-training and post-training errors for the beam stimulus.



| Precision | Accuracy | RMS |
|---|---|---|
| H(1,10) = 1.26, p = 0.2623 | H(1,10) = 5.03, p = 0.0250 | H(1,10) = 5.77, p = 0.0163 |

Another point of interest was whether learning would generalize to untrained target positions. Generalization could be tested only for those subjects who exhibited improvement. Kruskal-Wallis tests for the fine-grained testing stage, showed no significant difference between the subjects' performance on spatial locations for which they received training and for which they did not (see appendix B, Table 24). It is not possible to conclude that learning generalized to untrained target locations due to the lack of pre-training fine-grained testing data. However, the AVIT method can be claimed to have generated a continuous change of the auditory spatial map of TCD and TCE rather than a mere memorization of spatial locations. These generalization results support the hypothesis that the enhancement in sound localization ability ,due to AVIT, could potentially translate to real-world scenarios as opposed to just a well-controlled laboratory conditions.

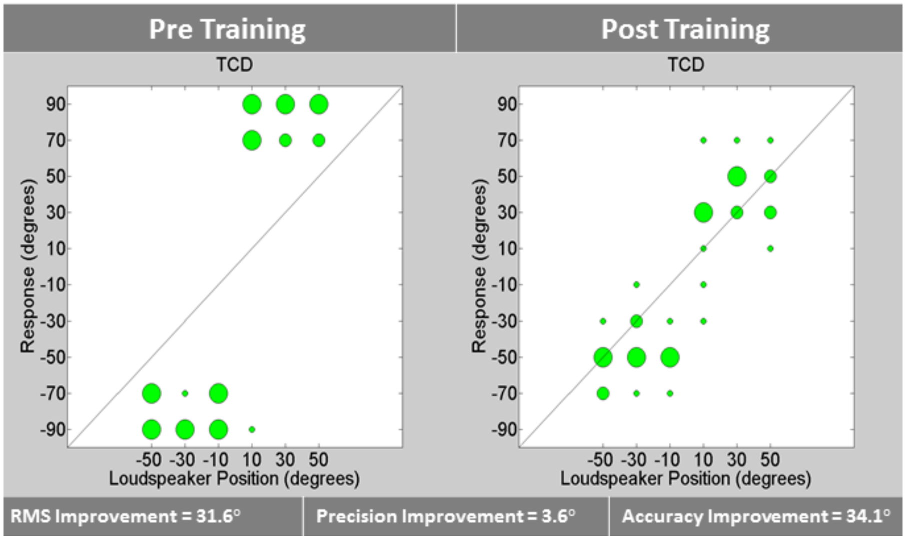


Figure 15. Pre-training and post-training localization data for subject TCD tested on the beam stimulus.

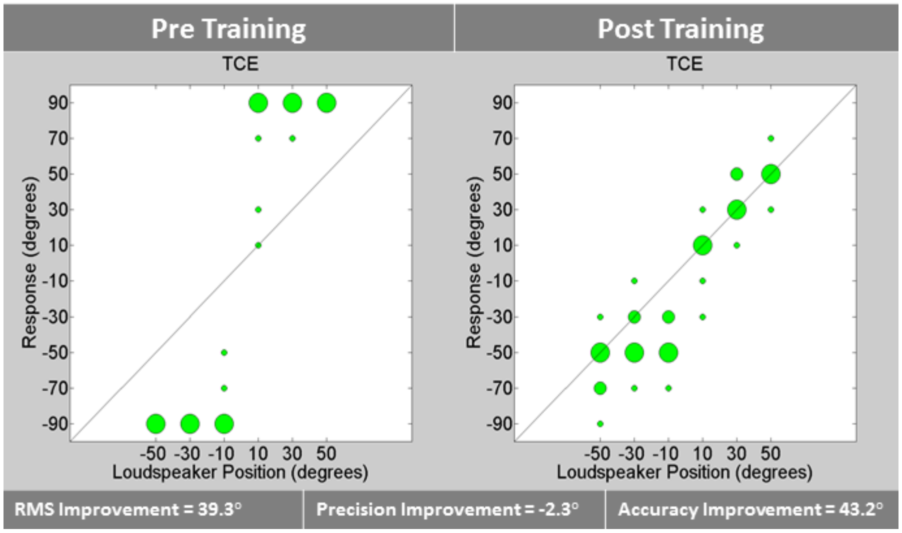


Figure 16. Pre-training and post-training localization data for subject TCE tested on the beam stimulus.

# IV. CONCLUSIONS

NH individuals were hypothesized to exhibit significant improvement in sound localization ability post AVIT, when tested in the two sub optimal listening environments in this study. This hypothesis is supported by the many ways in which multisensory training has been shown to facilitate learning. These include tasks like unilateral sound localization [31], detection, discrimination and memory tasks [65,66], visual learning and adaptation [67,68], visual motion detection [69], visual temporal order judgment [70], and auditory speech comprehension [44]. Multisensory training has even shown to increase learning rate [71].

In experiment 1 the specific hypothesis was that AVIT would enhance the precedence effect by increasing the echo threshold of listeners and thus allow them to better localize the leading sound. Bishop et al. (2011) support this hypothesis by showing that a visual stimulus coinciding with the leading sound enhances echo suppression. With the exception of subject SWL, the visually-trained listeners in experiment 1 exhibited a post-training localization performance ($E_{RMS} = 9.1° \pm 3.5°$) comparable to that of NH individuals localizing a single sound source, thereby confirming the hypothesis.

In experiment 2 the specific hypothesis was that AVIT would allow the listener to enhance their reliance on salient ILD cues and suppress their reliance on degraded ITD and spectral cues in order to improve localization performance. This hypothesis is supported by Sand and Nilsson (2014) who demonstrated that NH listeners could be trained to lateralize sound with only ILD cues available. Moreover, Strelnikov et al. (2011) showed that AVIT allowed listeners to enhance their reliance on salient spectral cues and suppress their reliance on degraded ILD and ITD cues in order to improve unilateral localization performance.

Contrary to the hypothesis, only two subjects in experiment 2 (TCD and TCE) exhibited significant improvement in localization ability of vocoded sounds. One possible reason for the difference in results between the Strelnikov et al. study and the present study is the consistency of the degraded cues. In the Strlenikov et al. study subjects had one of their ears plugged which degraded ILD and ITD cues but did so consistently across stimulus presentations. In the present study ITD and spectral cues were degraded but not consistently as there were 3 different versions of each stimulus corresponding to 3 different white noise carriers, and the stimulus in each presentation was randomly chosen from those three versions. Another difference in consistency is found in the level roving in the present study that was absent in the Strelnikov study. The less consistent stimuli in the present study may require more training in order to produce localization improvement across all subjects. Having said that, the less consistent stimuli made the training task more realistic which explains the generalization of localization improvement to an untrained stimulus exhibited by subjects TCD and TCE [49].

One possible explanation for the stark difference between the two subjects in the visually-trained group that improved significantly (TCD and TCE) and the other subjects in that group that did not, can be found in Wang et al. (2008). Wang et al. tested two groups of NH listeners on a speech identification task. One of the conditions of the task involved Mcgurk effect pairs, which are one auditory syllable and one visual syllable (video of lip movement) that when integrated result in the perception of a third distinct syllable. Both groups consisted of Chinese born adults that moved to Canada. Their native language was Mandarin and English was their second language. One group resided in Canada for 2 years at the time of the study and the other for 10 years. The results showed that the group that was exposed to the English language for the shorter amount of time, performed better on the Mcgurk pair condition. Thus, individuals that reside in a foreign language country may possess higher ability to integrate auditory and visual cues due to their partial reliance on lip reading. Subjects TCD and TCE were the only subjects in the visually-trained group that fit this description. Moreover, the top two performers in the visually-trained group in experiment 1 (SWA and SWB) weren’t native English speakers as well.

The Wang et al. study thus also supports the assumption that AVIT would be successful with CI users as they partially rely on lip reading as well.

The high variability in localization improvement between subjects, including some subjects that showed reduced localization ability post-training, could also be a result of varying attention levels between subjects. Attention has been shown to be a significant factor in sound localization (Bergan et al., 2005; Teder-Sälejärvi et al., 1999; Teder-Sälejärvi and Hillyard, 1998). Moreover, the long testing days combined with the task's inherent difficulty made the need for attention that much more significant. This suggests that the amount of training necessary to generate supervised learning could be subject-dependent, as the capacity for prolonged attention is subject-dependent as well.

Perhaps a more extensive training paradigm is needed to show that AVIT can generate supervised learning in CI simulated environments. This could mean a more scaffolded approach to training, in which the procedure is split up to more than 3 days, in order to decrease the time requirement for continuous attention. The added study time would make subject recruitment more challenging but could also allow for visual reinforcement on a larger number of loudspeakers for a longer duration of time.

AVIT has been used as a rehabilitation tool for individuals with visual hemineglect and hemianopsia [76,77]. Strelnikov et al. (2011) have suggested that AVIT could be used to rehabilitated individuals with unilateral deafness or with a single CI. Although further testing is needed (with a longer or more frequent training procedure), this study suggests that AVIT could also be used to rehabilitate individuals with bilateral cochlear implants.

# APPENDIX A: ERROR TYPES

## RMS Error

$$E_{RMS} = \sqrt{\frac{\sum_{i=1}^{S}\sum_{j=1}^{R}\left(\theta_{i,target}-\theta_{i,j,response}\right)^2}{S*R}}$$

Where S is the number of loudspeakers, R is the number of times the stimulus is presented at each loudspeaker (Repetitions), $\theta_{i,target}$ is the loudspeaker’s position in degrees azimuth, and $\theta_{i,j,response}$ is the perceived position of the sound for the i-th loudspeaker and j-th repetition.

## Variability Error

$$E_{Var} = \sqrt{\frac{\sum_{i=1}^{S}\sum_{j=1}^{R}\left(\theta_{avg\_i,response} - \theta_{i,j,response}\right)^2}{S * R}}$$

Where $\theta_{avg\_i,response} = \sum_{j=1}^{R}\frac{\theta_{i,j,response}}{R}$

## Bias Error

$$E_{Bias} = \sqrt{\frac{\sum_{i=1}^{S}\left(\theta_{i,target} - \theta_{avg\_i,response}\right)^2}{S}}$$

# APPENDIX B: SUPPLAMENTAL STATISTICAL DATA

Table 21. Kruskal-Wallis statistics comparing pre-training and post-training errors for the white noise burst stimulus.

| Precision | Accuracy | RMS |
|---|---|---|
| H(1,10) = 0.16, p = 0.6884 | H(1,10) = 0.41, p = 0.5218 | H(1,10) = 0.64, p = 0.4233 |

Table 22. Kruskal-Wallis statistics comparing error improvement between the visually-trained and control groups for the beam stimulus.

| Accuracy Improvement | Normalized Accuracy Improvement | RMS Improvement | Normalized RMS Improvement |
|---|---|---|---|
| H(1,9) = 0.53<br>p = 0.4652 | H(1,9) = 1.41<br>p = 0.2343 | H(1,9) = 1.63<br>p = 0.2012 | H(1,9) = 3.02<br>p = 0.0821 |

Table 23. Kruskal-Wallis statistics comparing performance (in terms of RMS errors) between trained target locations and untrained target locations in the echoic environment.

| | SVS | SVT | SWA | SWB | SWN |
|---|---|---|---|---|---|
| Trained stimulus | H(1,10) = 0.16<br>p = 0.6884 | H(1,10) = 3.71<br>p = 0.0542 | H(1,10) = 0.1<br>p = 0.7488 | H(1,10) = 1.26<br>p = 0.2615 | H(1,10) = 0.23<br>p = 0.6298 |

Table 24. Kruskal-Wallis statistics comparing performance (in terms of RMS errors) between trained target locations and untrained target locations in the CI simulated environment.

| | TCD | TCE |
|---|---|---|
| Trained stimulus | H(1,10) = 0.64<br>p = 0.4233 | H(1,10) = 3.4<br>p = 0.0651 |
| Untrained stimulus | H(1,10) = 0.03<br>p = 0.8728 | H(1,10) = 1.65<br>p = 0.1994 |

# APPENDIX C: SUPPLAMENTAL LOCALIZATION DATA

## Progress data for sine tone stimulus

The data below describes the progress of each subject through the 3 study days. The top three rows describe the subject's progress on stage 1, stage 2 and stage 3 respectively. The bottom row describes the subject's performance on the single source task, pre-training task, post-training task and the fine-grained task. The pre- and post-training data are highlighted in magenta for the reader's convenience. Note that the fourth testing block on the third stage constitutes the post-training task and thus their data is identical.

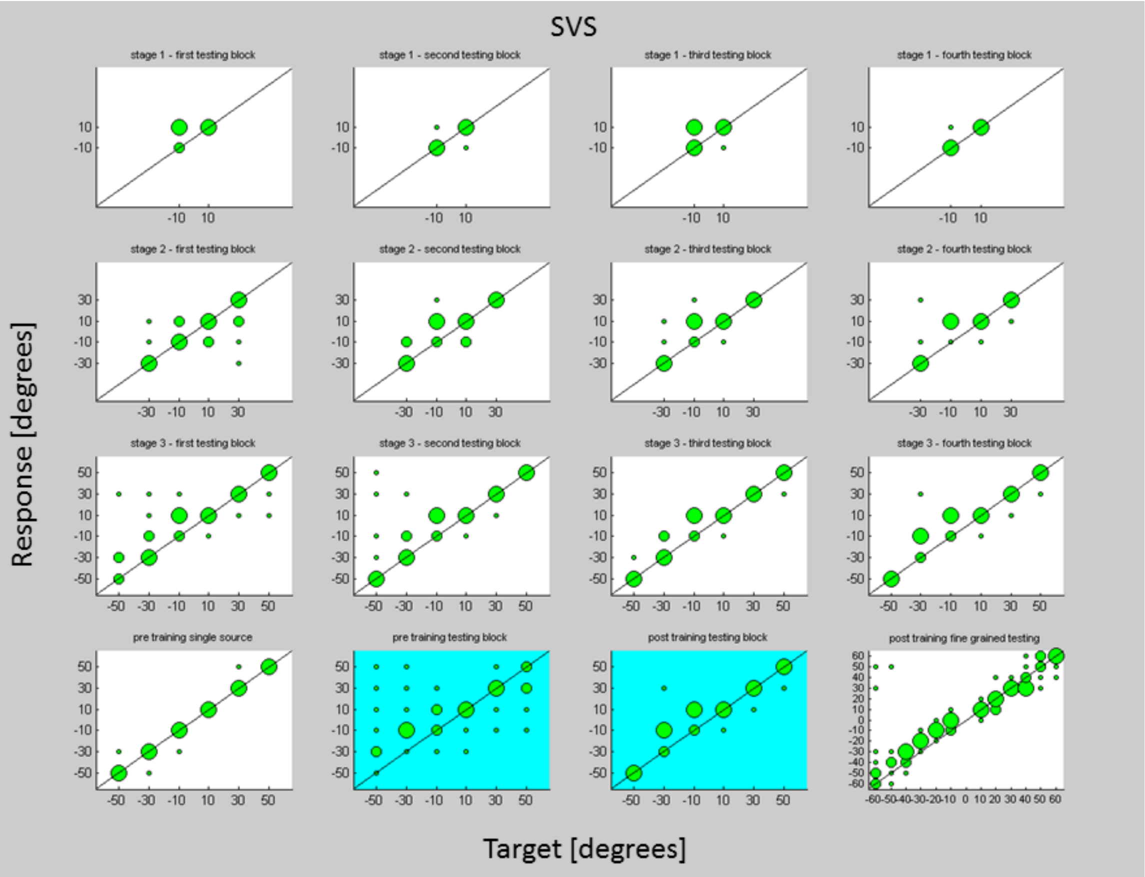


Figure 17

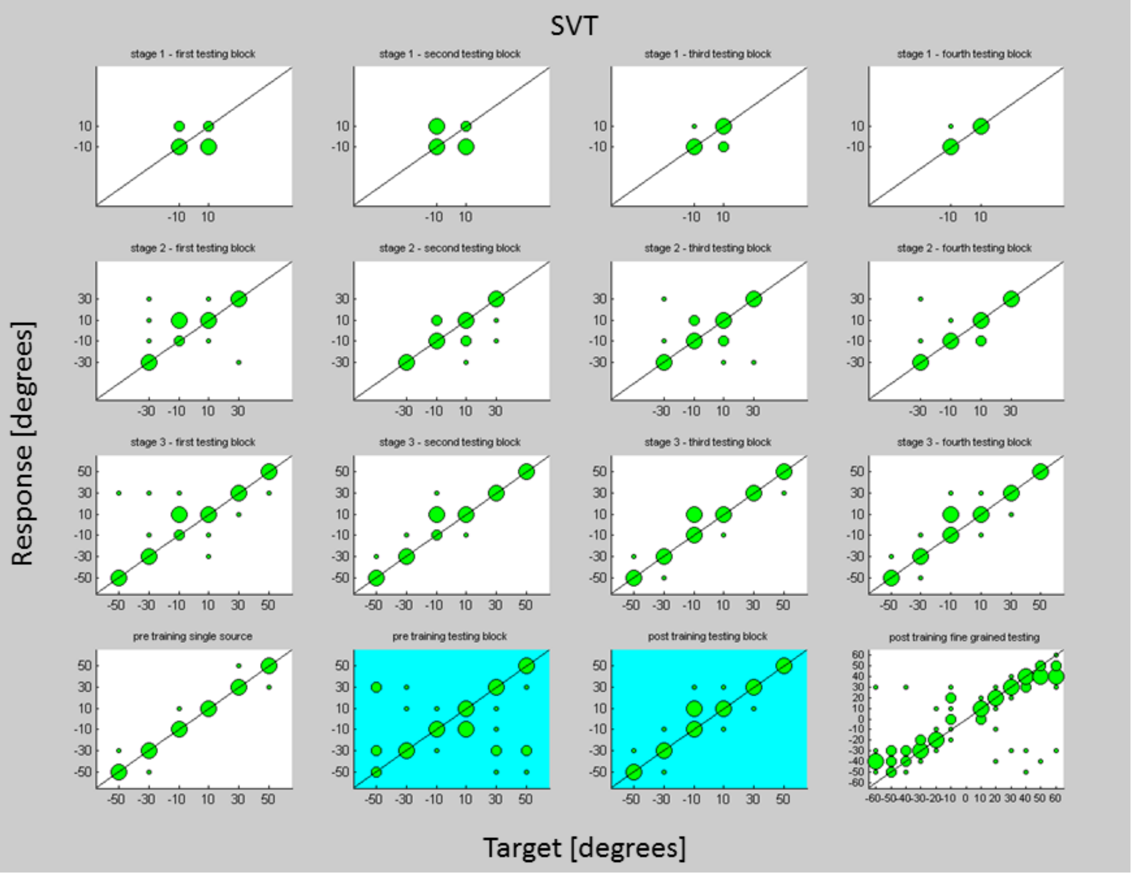

SVT
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
post training fine grained testing
Response [degrees]
Target [degrees]
-50 -30 -10 10 30 50
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60

Figure 18

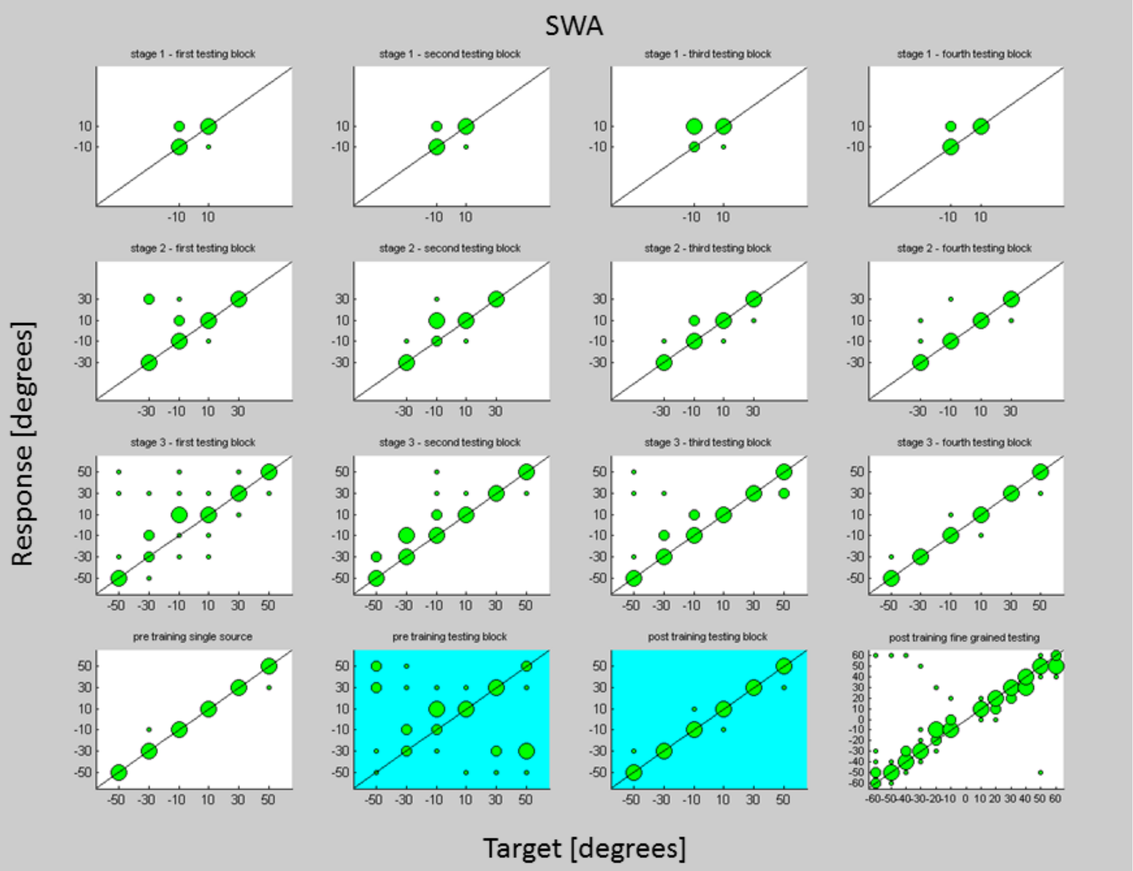

SWA
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
post training fine grained testing
-50 -30 -10 10 30 50
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 19

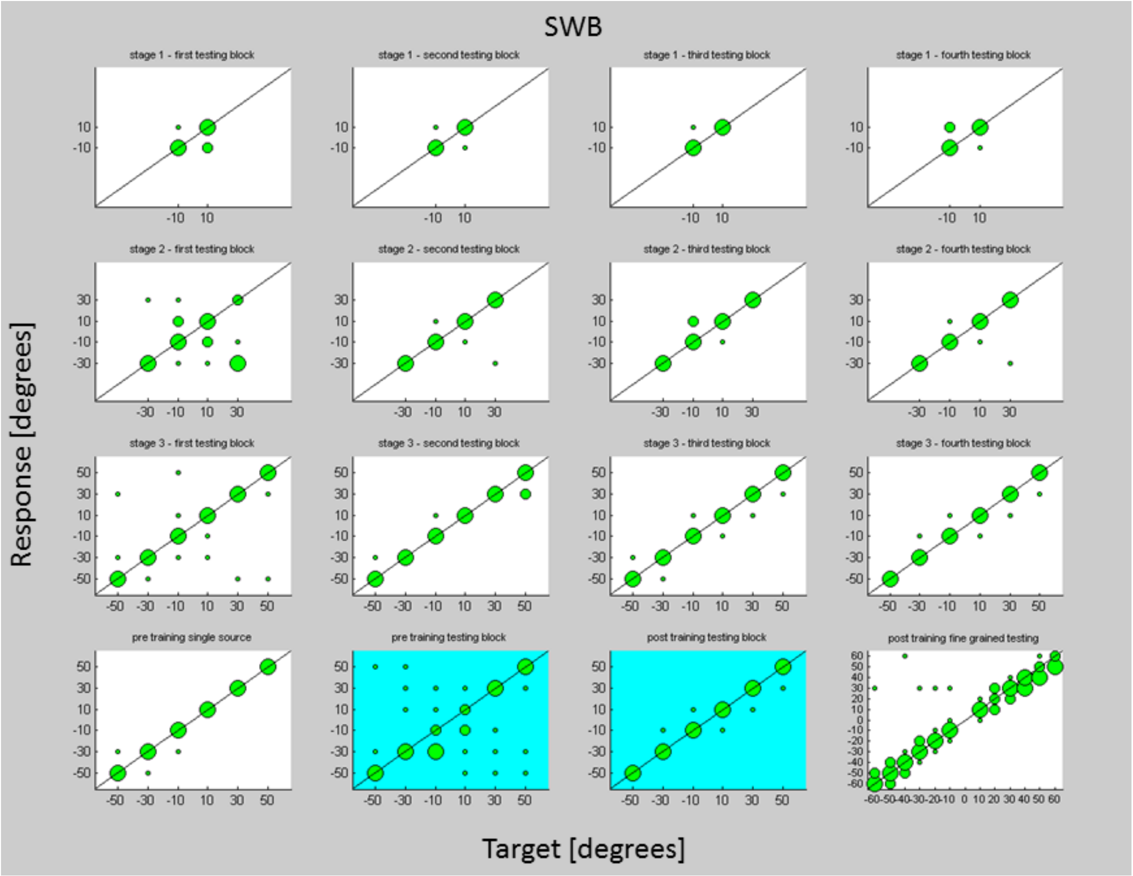

SWB
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
post training fine grained testing
10
-10
30
-30
50
-50
-60-50-40-30-20-10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 20

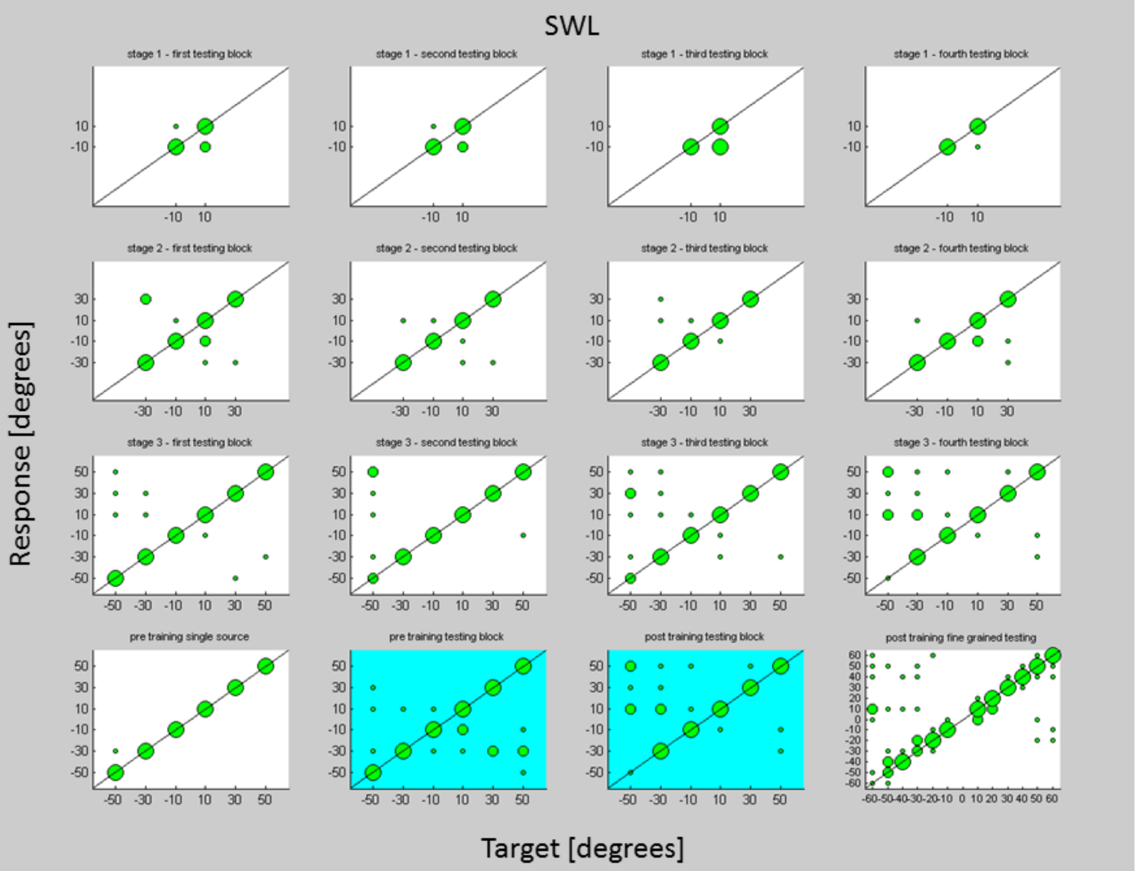

SWL
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
10
-10
-10 10
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
30
10
-10
-30
-30 -10 10 30
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50
30
10
-10
-30
-50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
post training fine grained testing
60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60
-60-50-40-30-20-10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 21

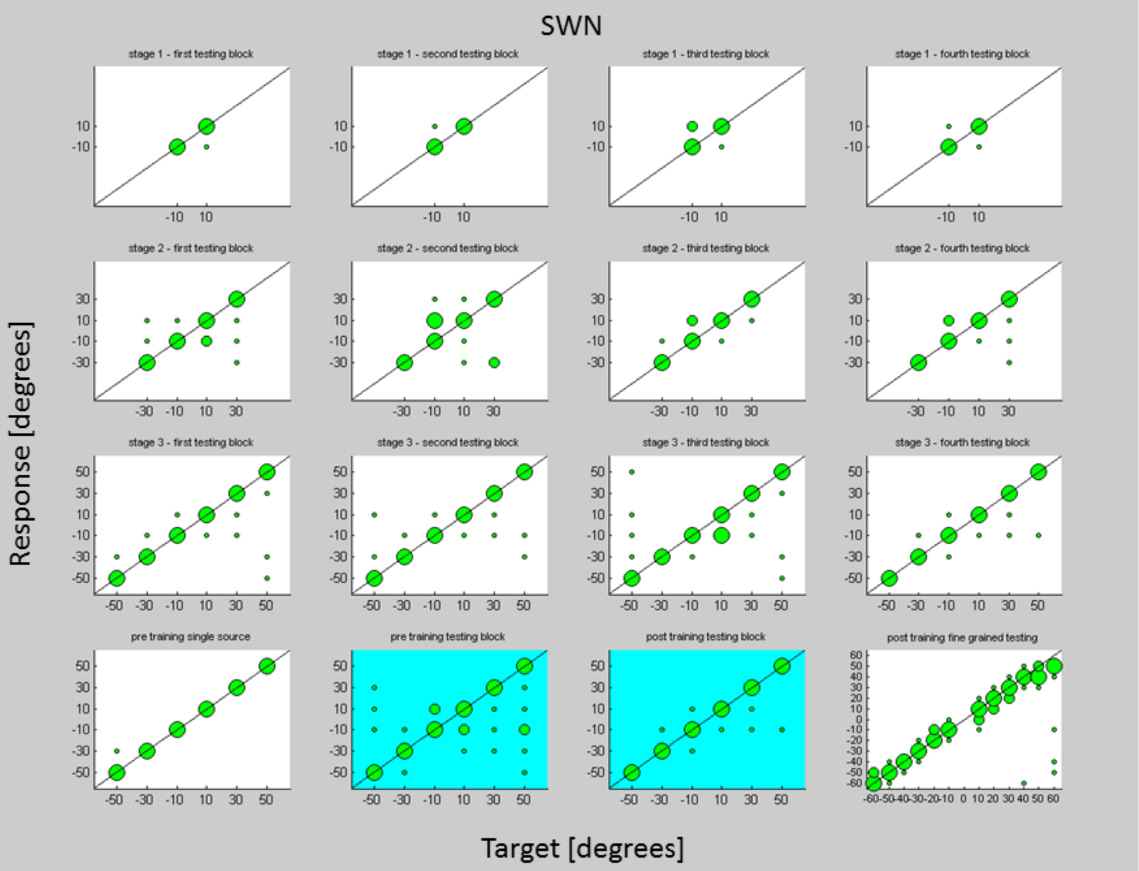

SWN
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
post training fine grained testing
-50 -30 -10 10 30 50
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 22

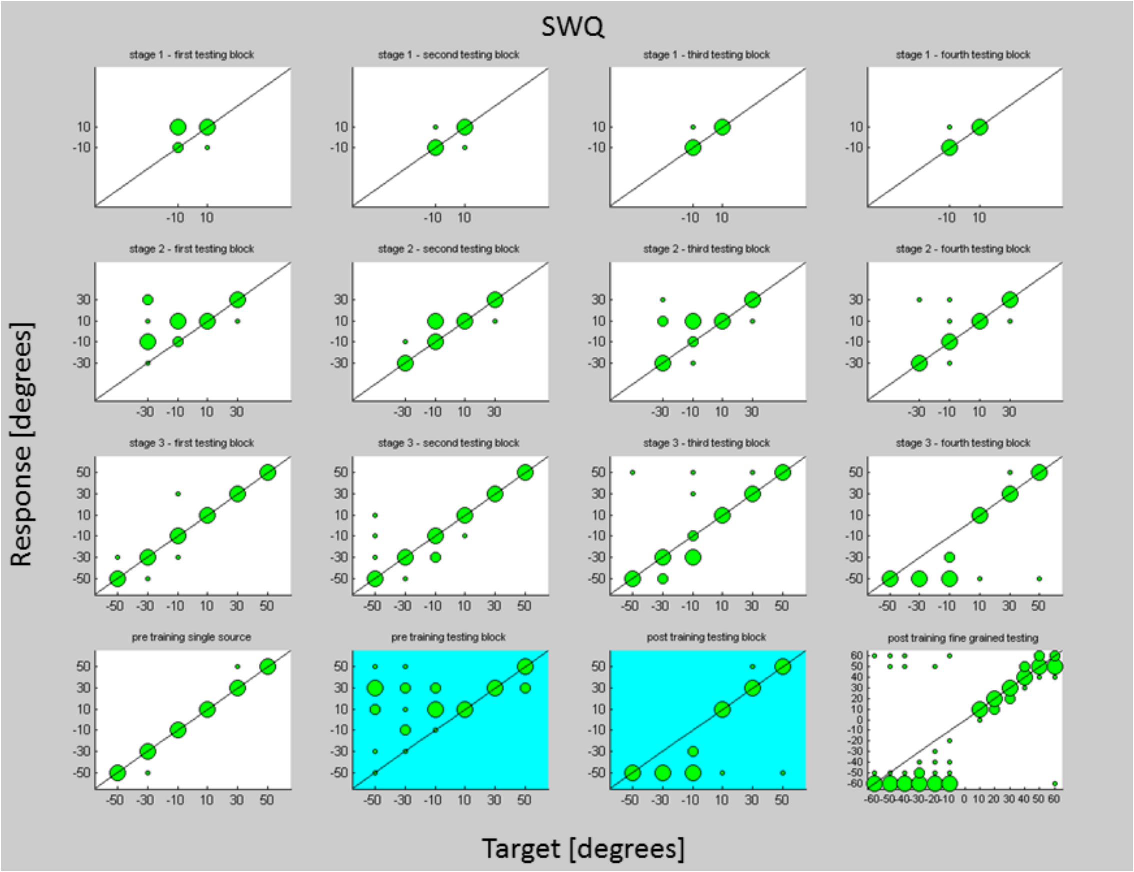

SWQ
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
10
-10
-10 10
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
30
10
-10
-30
-30 -10 10 30
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50
30
10
-10
-30
-50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
post training fine grained testing
60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 23

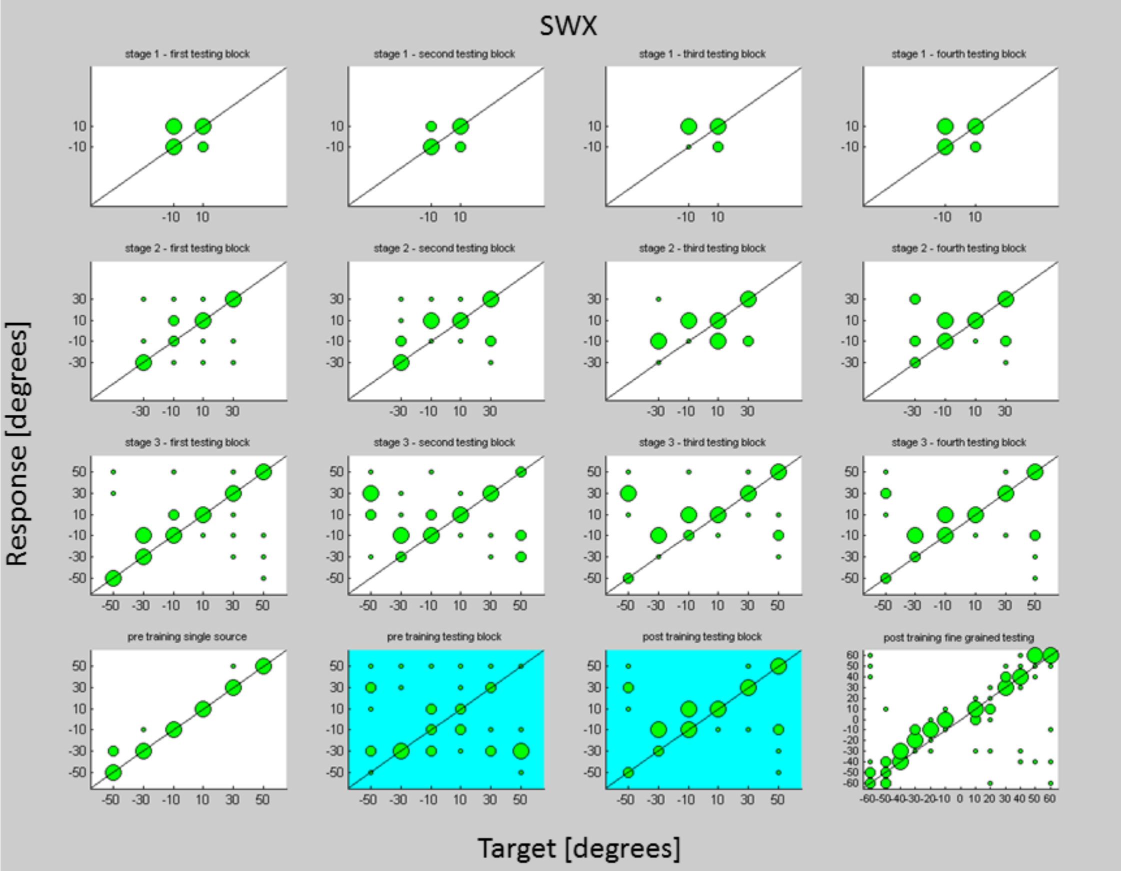

SWX
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
post training fine grained testing
Response [degrees]
Target [degrees]
-50 -30 -10 10 30 50
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60

Figure 24

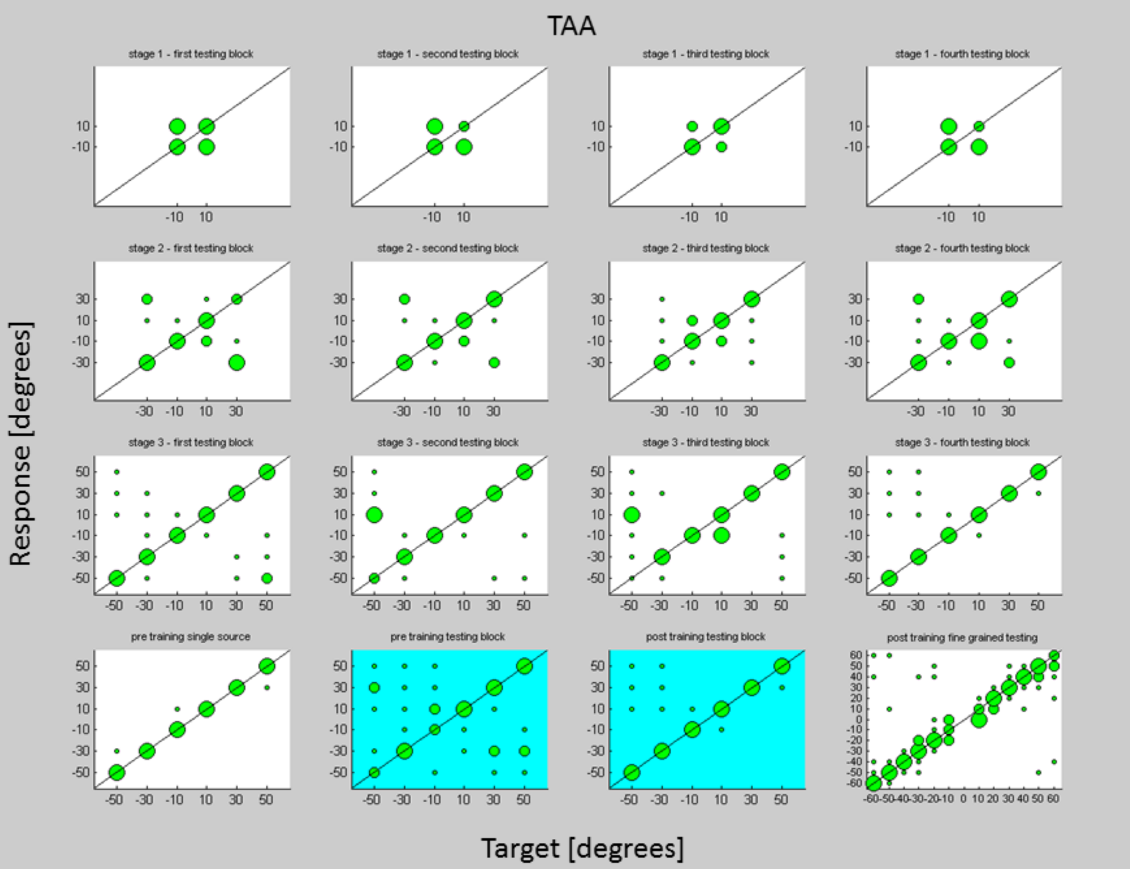

TAA
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
post training fine grained testing
10
-10
30
50
-30
-50
-60-50-40-30-20-10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 25

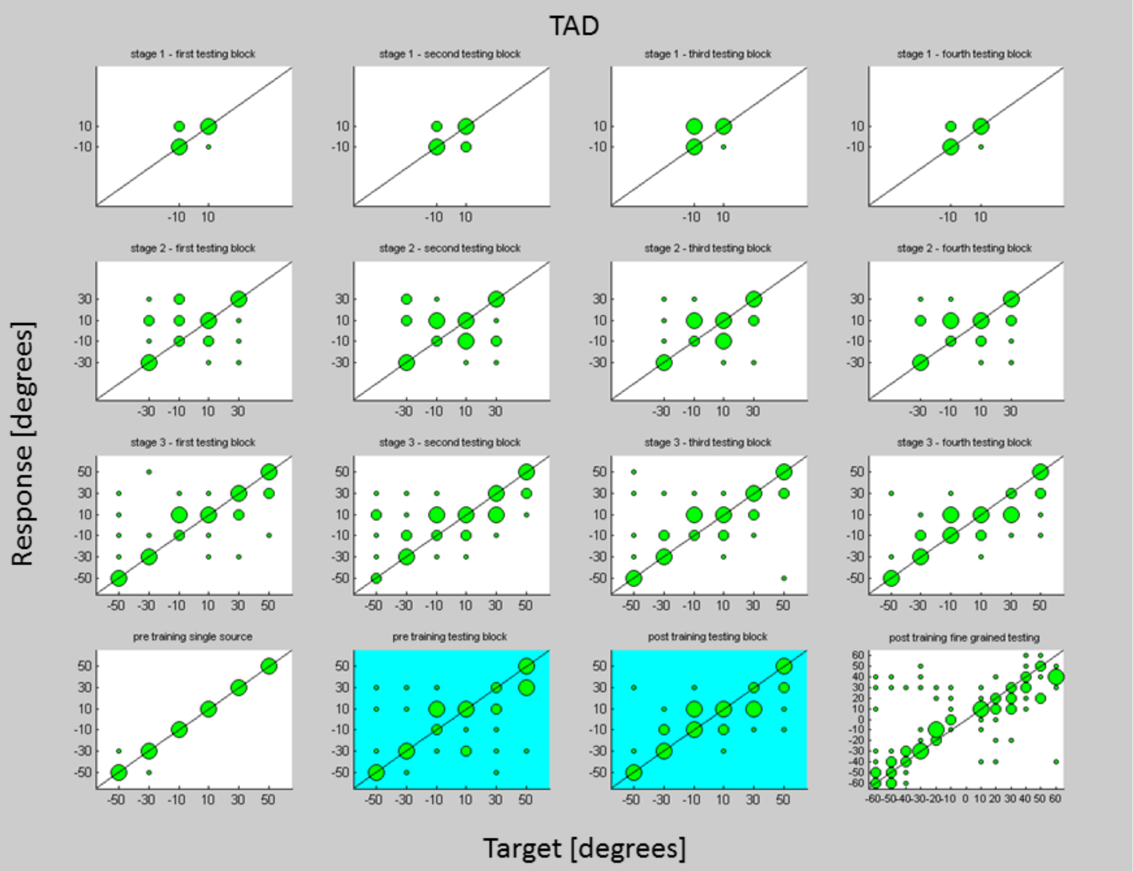

TAD
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
10
-10
-10 10
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
30
10
-10
-30
-30 -10 10 30
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50
30
10
-10
-30
-50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
post training fine grained testing
60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 26

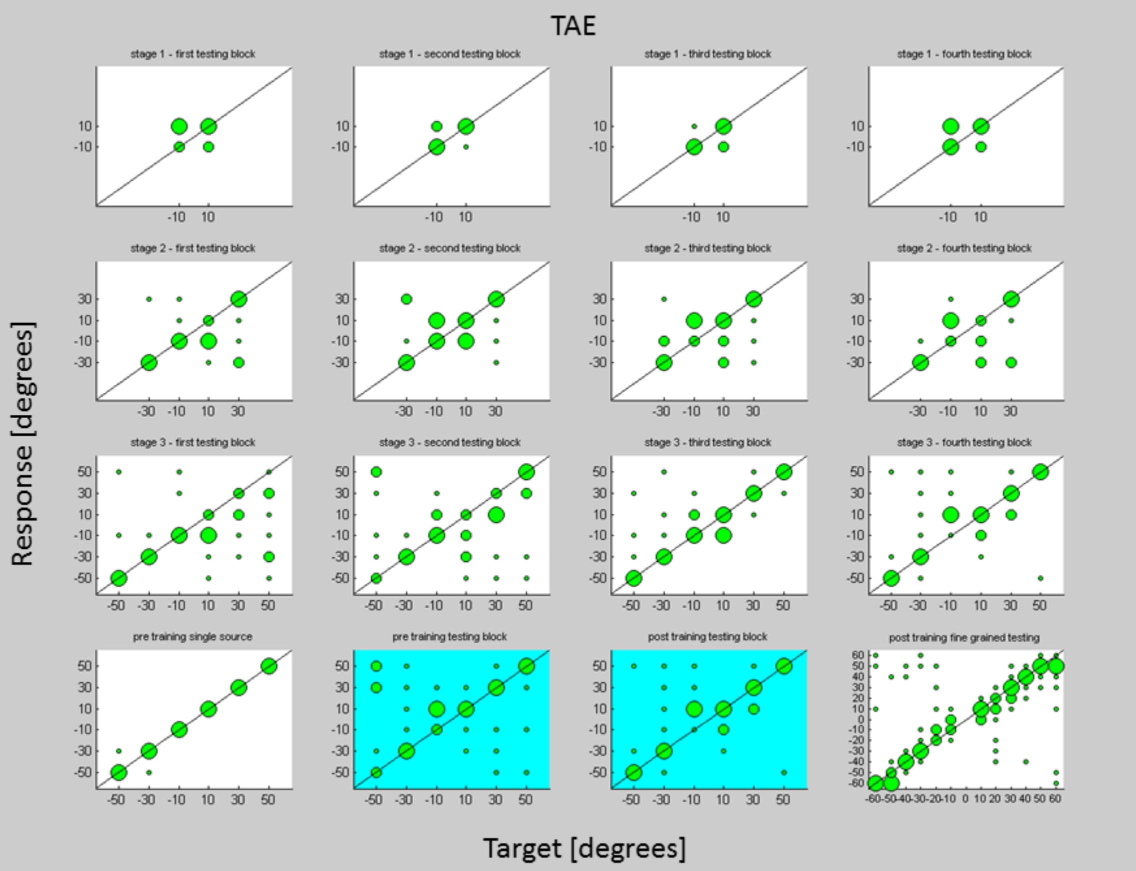

TAE
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
10
-10
-10 10
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
30
10
-10
-30
-30 -10 10 30
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50
30
10
-10
-30
-50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
post training fine grained testing
60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60
-60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60
Response [degrees]
Target [degrees]

Figure 27

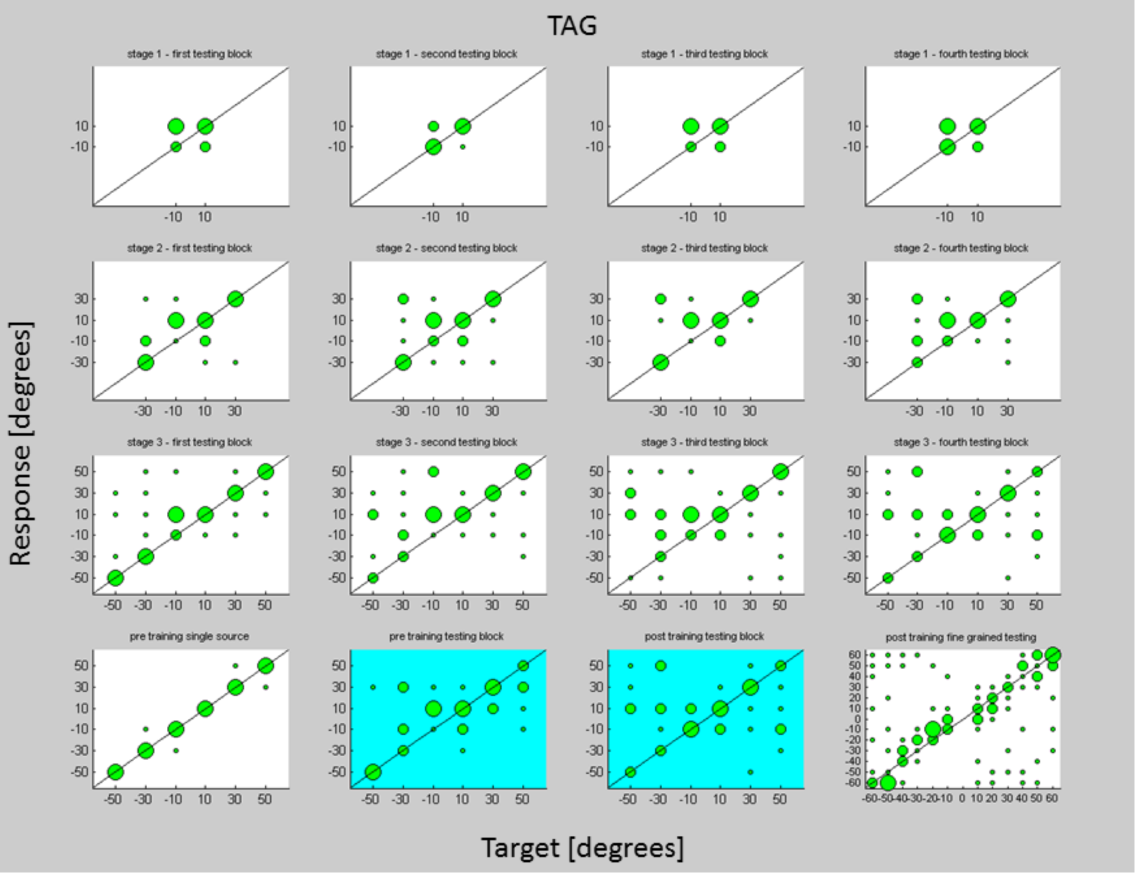


Figure 28

## Data for white noise burst stimulus

The data below describes the subjects' performance on the pre-training task, post-training task and the fine-grained task. The white noise burst is the stimulus that the subjects were not trained on, thus there is no other "progress" data.

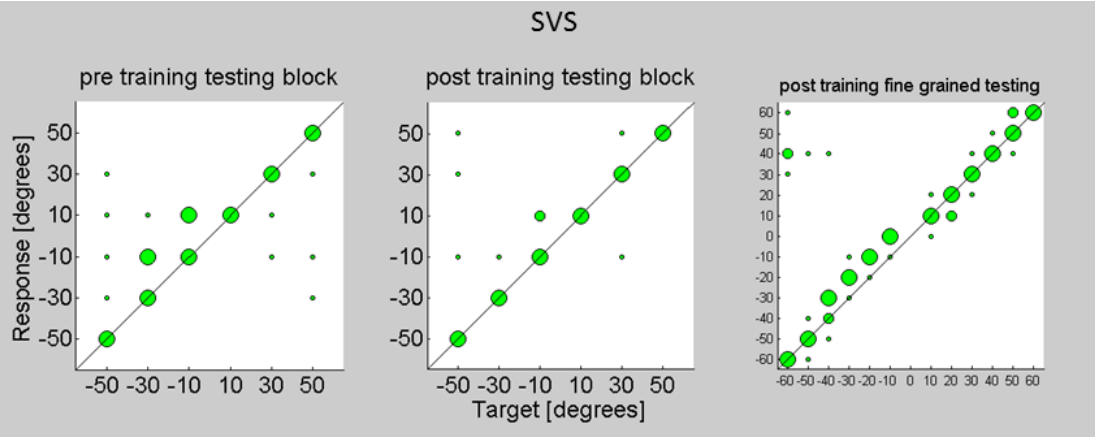


Figure 29

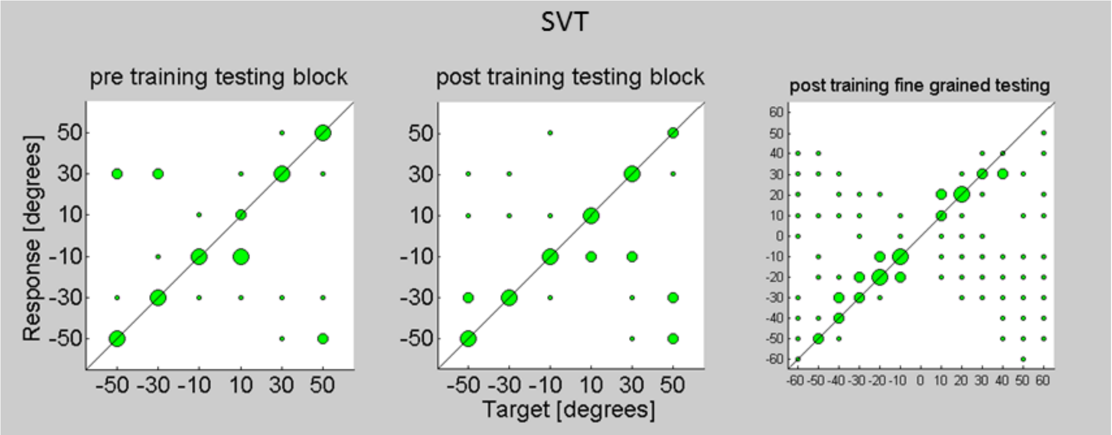


Figure 30

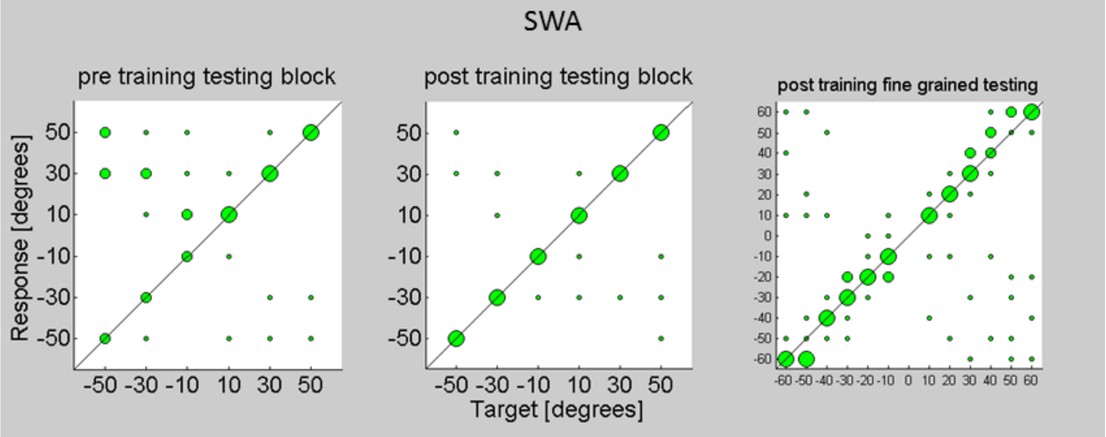


Figure 31

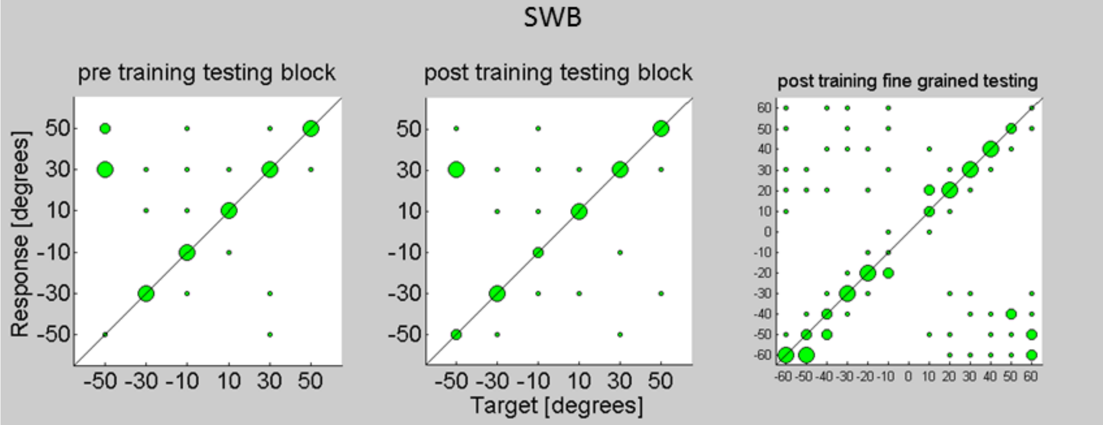


Figure 32

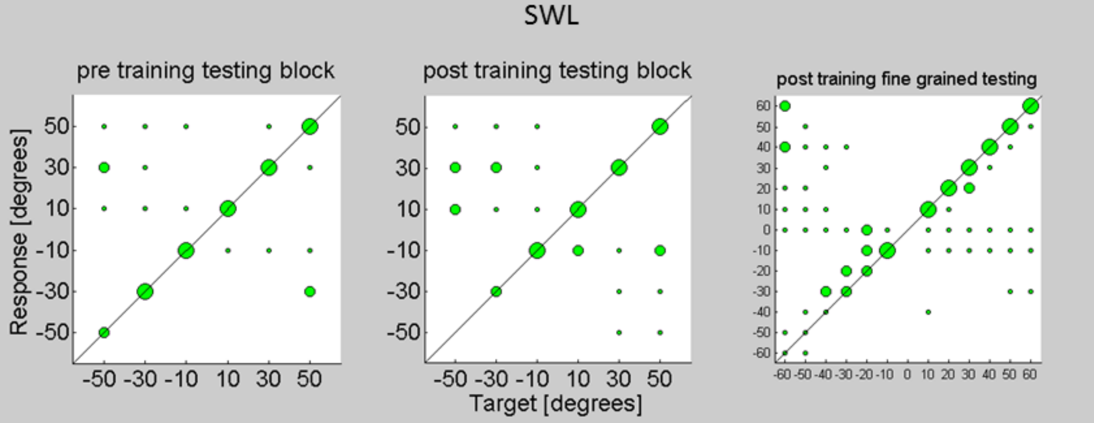


Figure 33

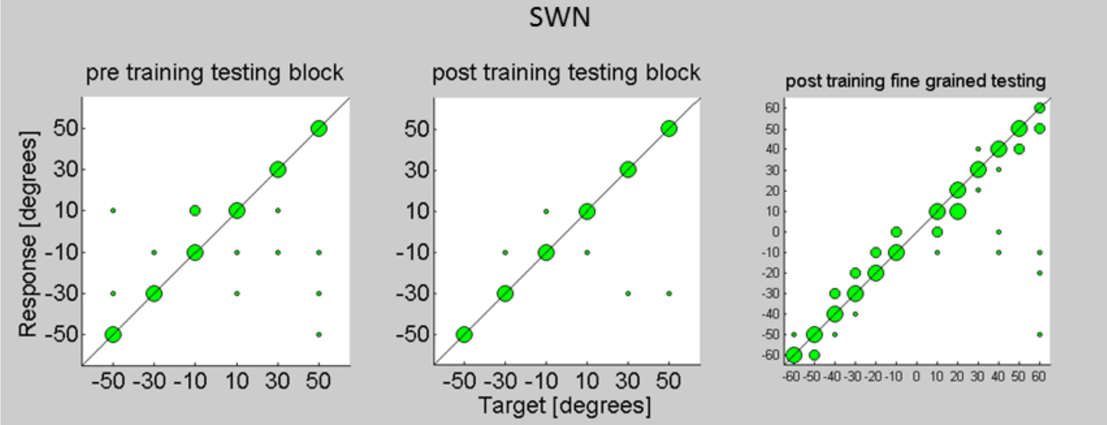


Figure 34

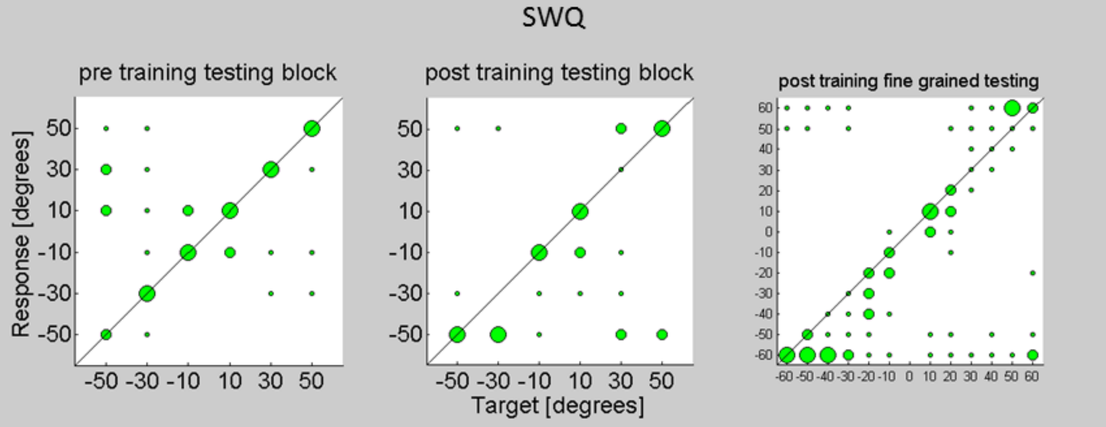


Figure 35

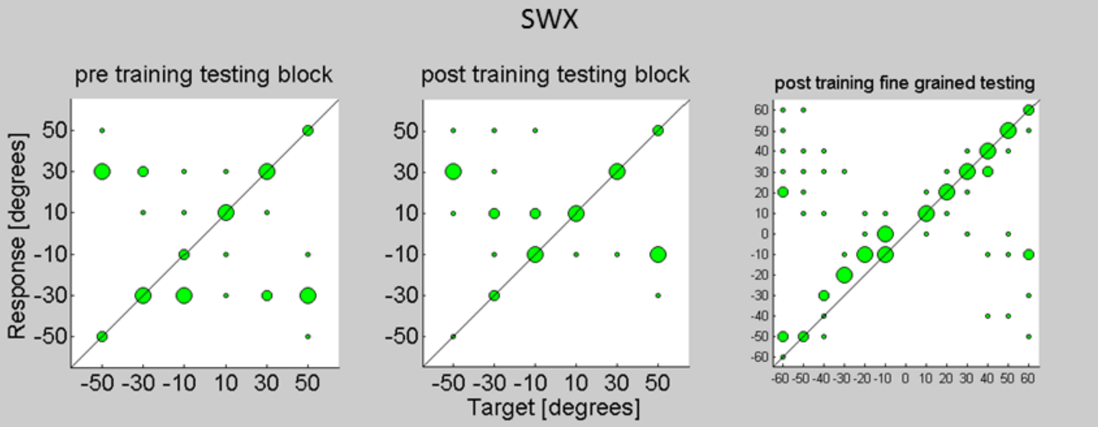


Figure 36

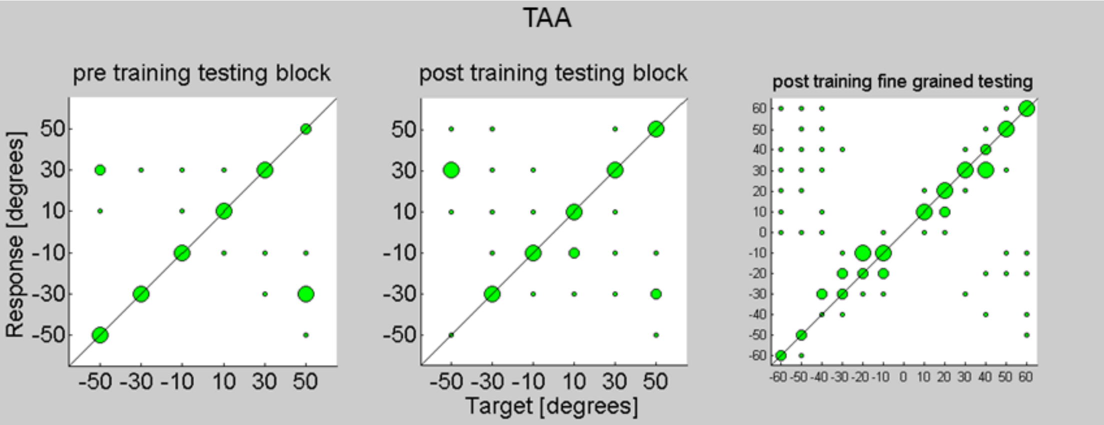


Figure 37

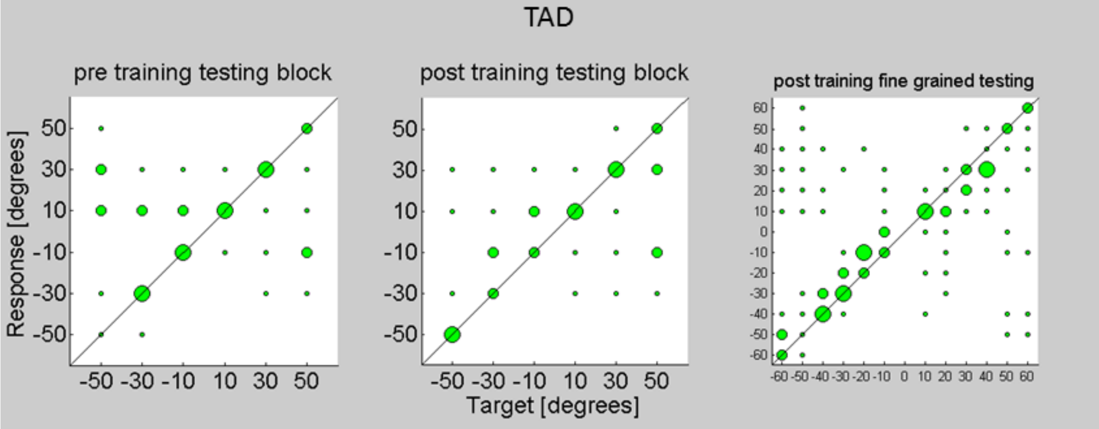


Figure 38

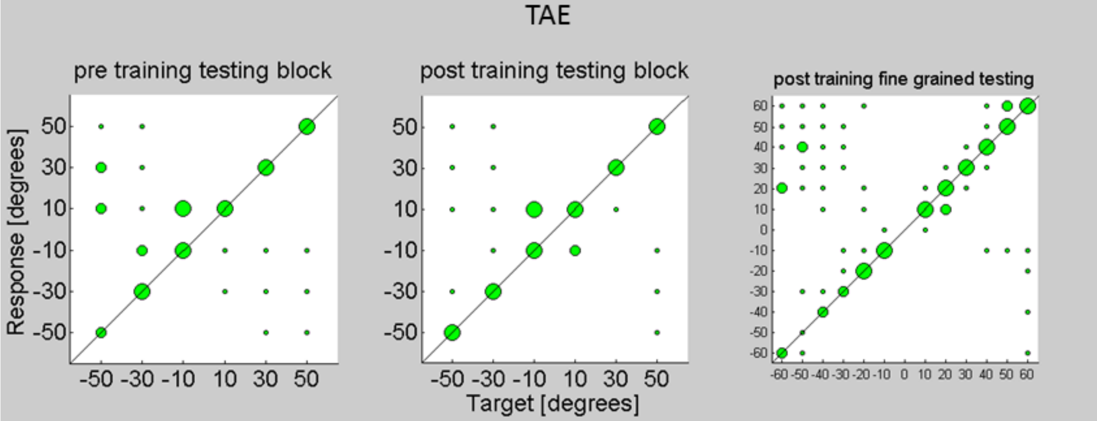


Figure 39

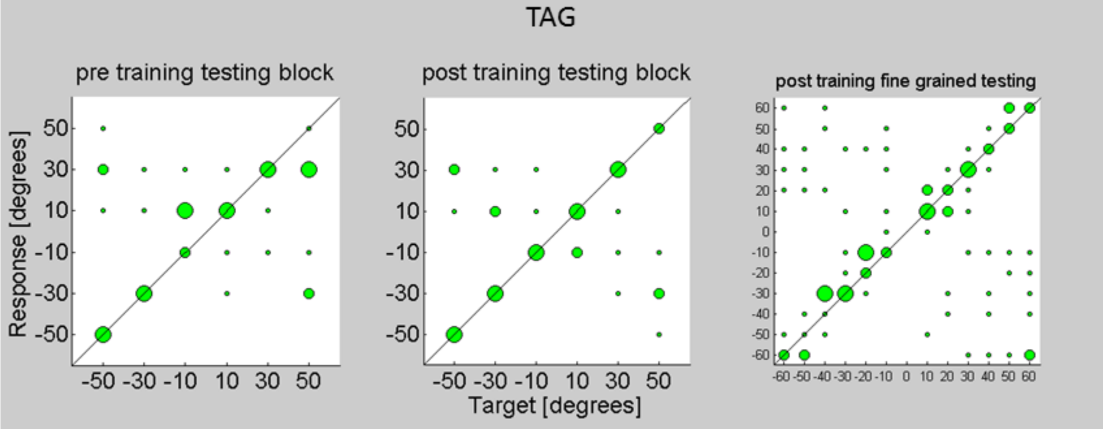


Figure 40

### Progress data for 'merge' stimulus

The data below describes the progress of each subject through the 3 study days. The top three rows describe the subject's progress on stage 1, stage 2 and stage 3 respectively. The bottom row describes the subject's performance on the single source task, pre-training task and post-training task. The pre- and post-training data are highlighted in magenta for the reader's convenience. The fine-grained testing data is presented separately for the visually-trained group and for the control group.

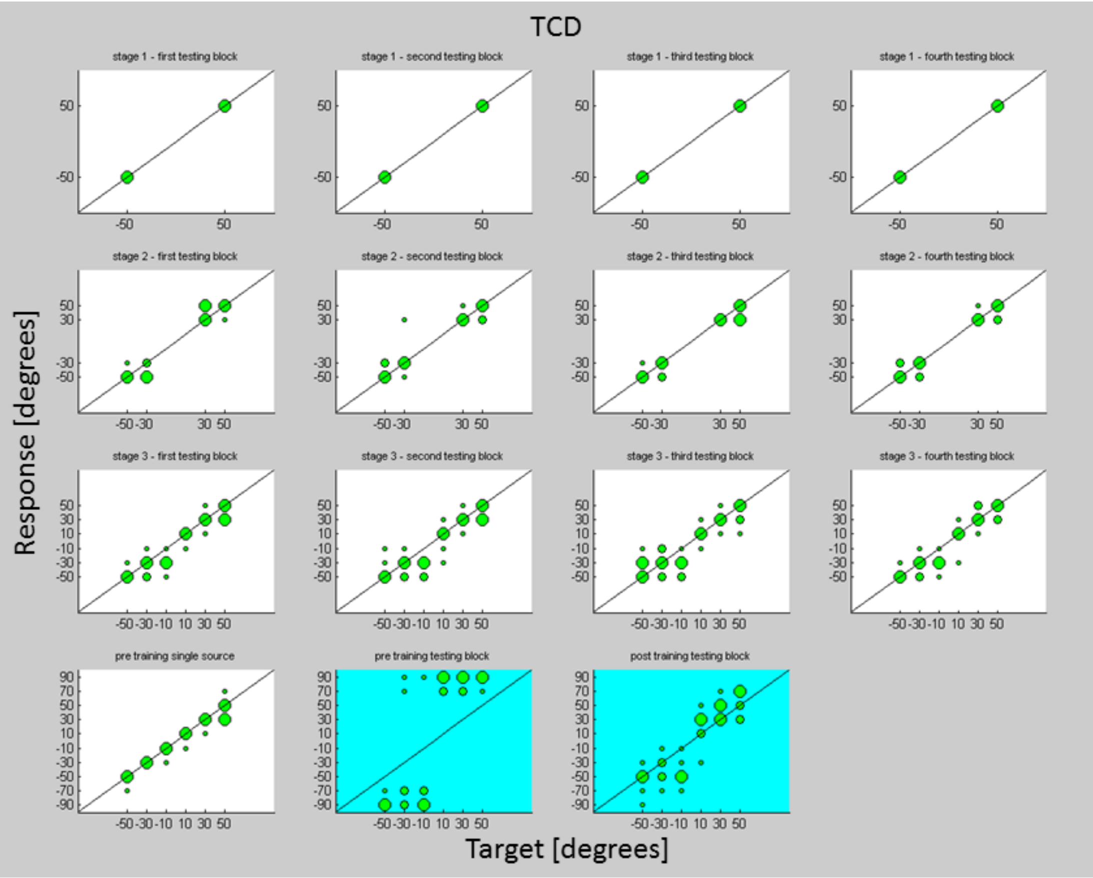


Figure 41

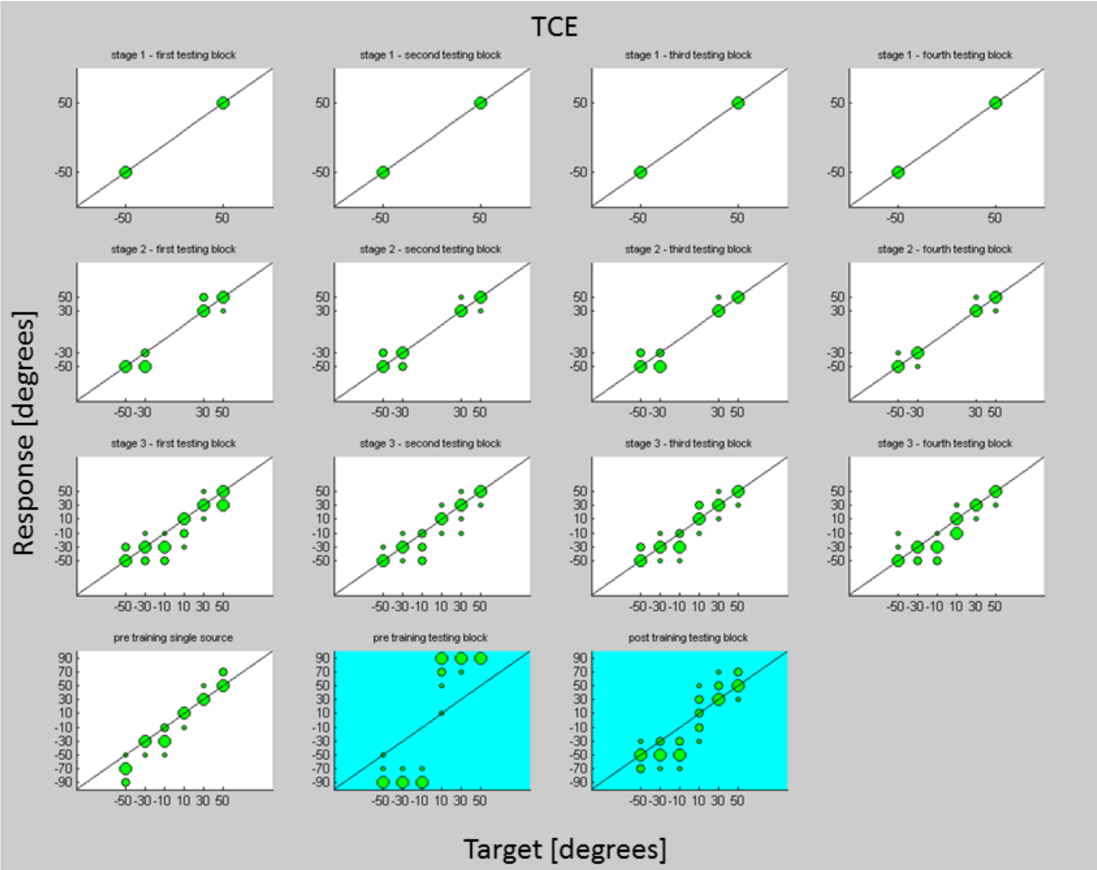

TCE
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
50
-50
-50
50
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
50
30
-30
-50
-50 -30
30 50
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50
30
10
-10
-30
-50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
Response [degrees]
Target [degrees]

Figure 42

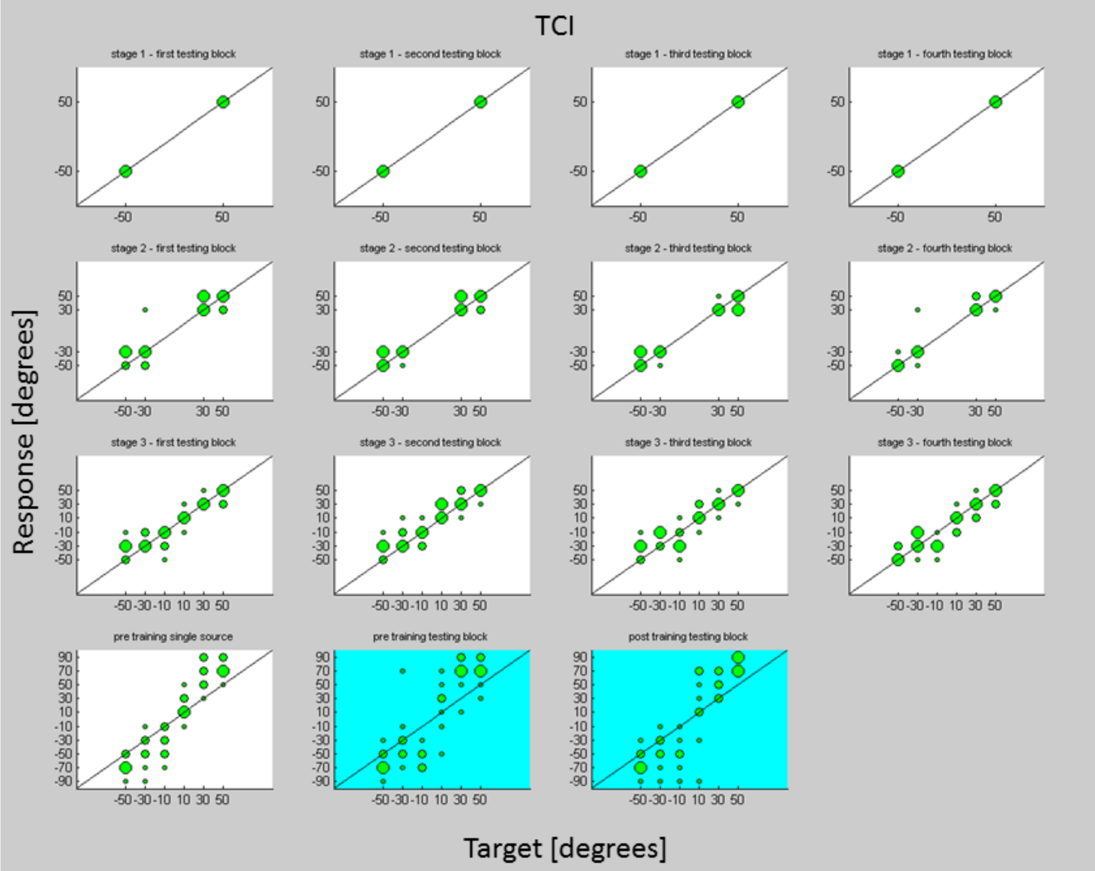

TCI
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
50
-50
-50
50
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
50
30
-30
-50
-50 -30
30 50
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50
30
10
-10
-30
-50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
Response [degrees]
Target [degrees]

Figure 43

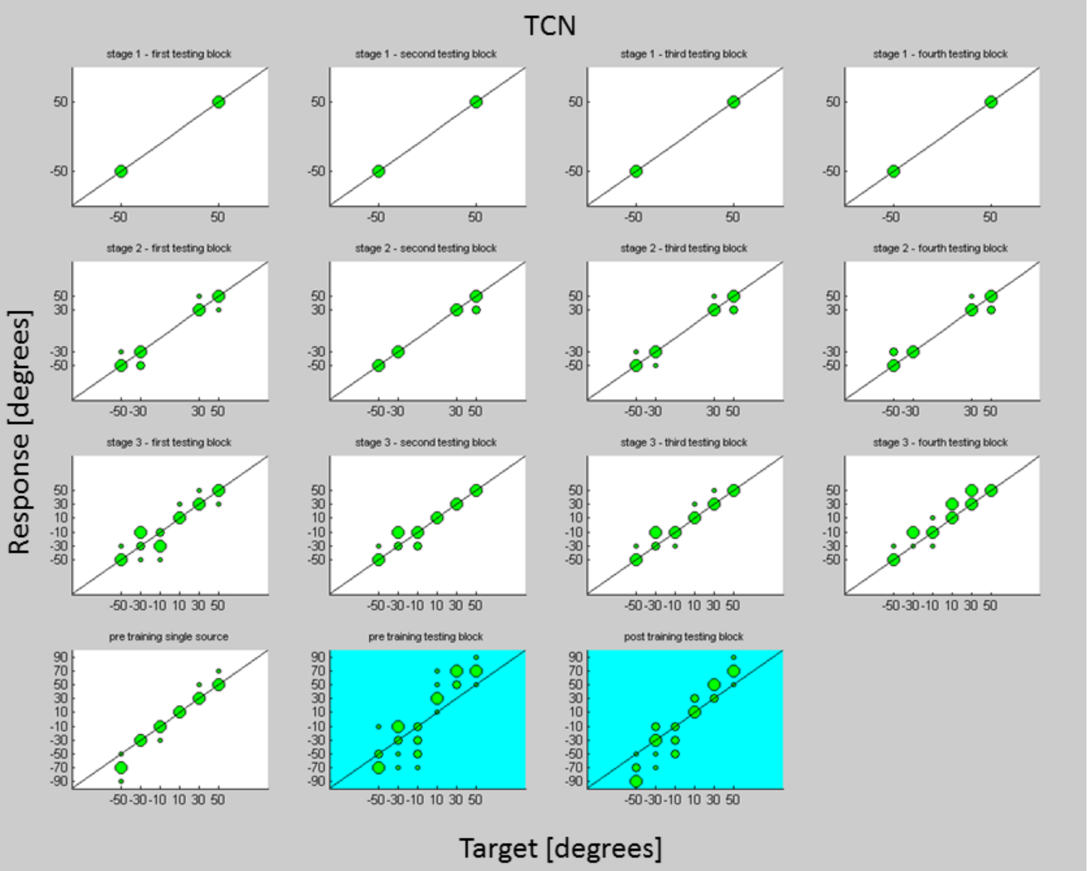

TCN
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
50
-50
-50 50
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
50 30 -30 -50
-50 -30 30 50
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50 30 10 -10 -30 -50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
90 70 50 30 10 -10 -30 -50 -70 -90
-50 -30 -10 10 30 50
Response [degrees]
Target [degrees]

Figure 44

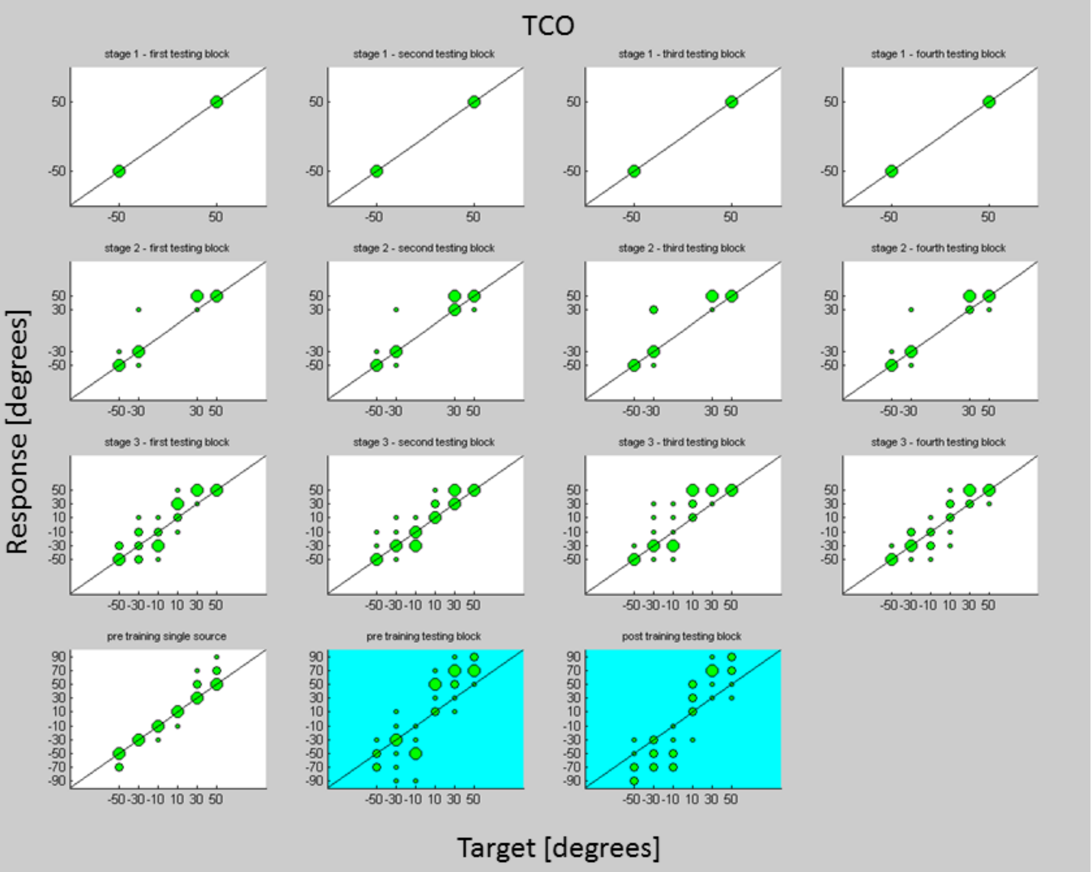

TCO
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
50
-50
-50 50
50 30 -30 -50
-50 -30 30 50
50 30 10 -10 -30 -50
-50 -30 -10 10 30 50
90 70 50 30 10 -10 -30 -50 -70 -90
Response [degrees]
Target [degrees]

Figure 45

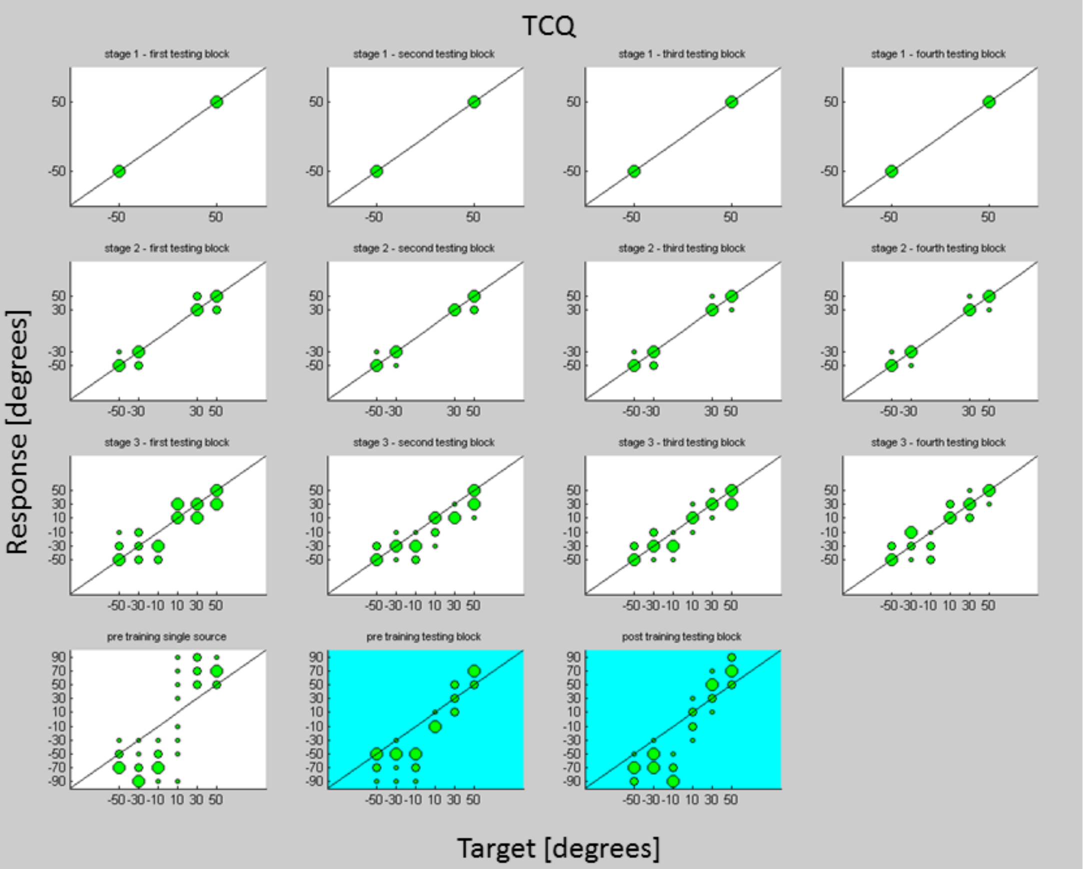

TCQ
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
-50
50
-30
30
-10
10
70
90
-70
-90
Response [degrees]
Target [degrees]

Figure 46

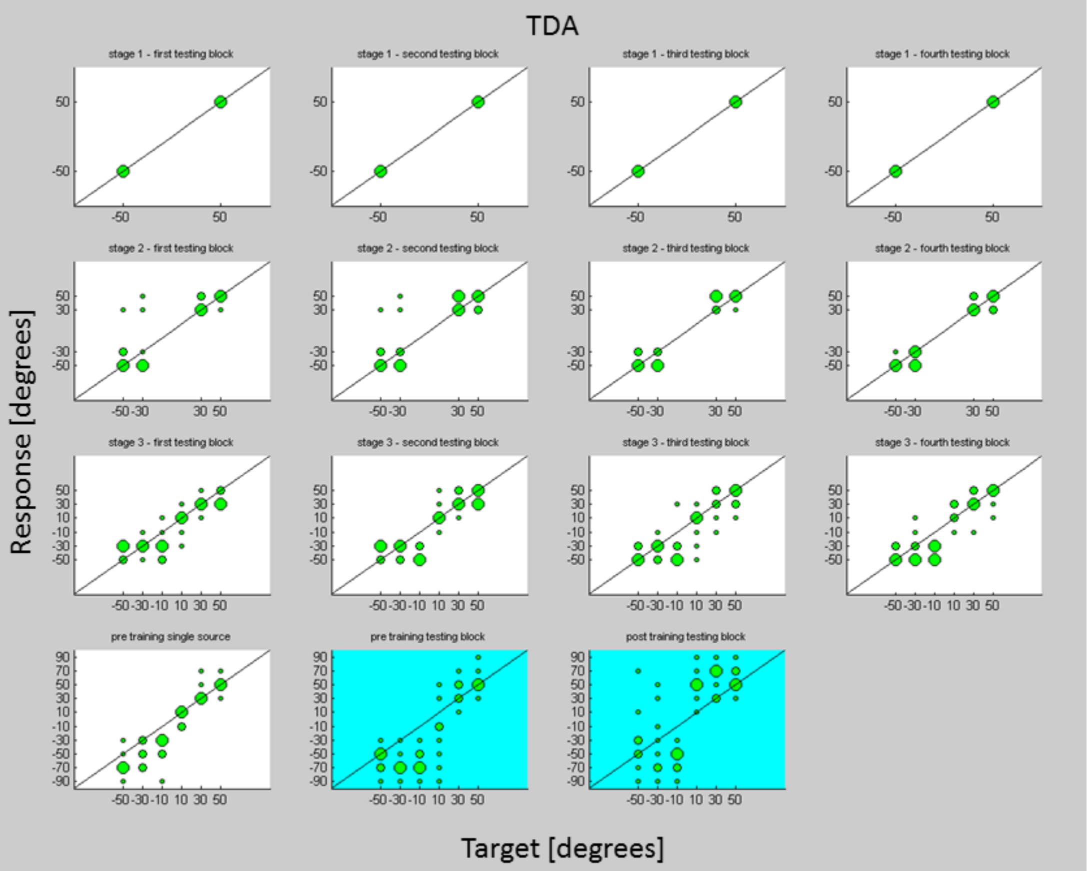

TDA
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
50
-50
30
-30
10
-10
90
70
-70
-90
-50
50
-50 -30
30 50
-50 -30 -10 10 30 50
Response [degrees]
Target [degrees]

Figure 47

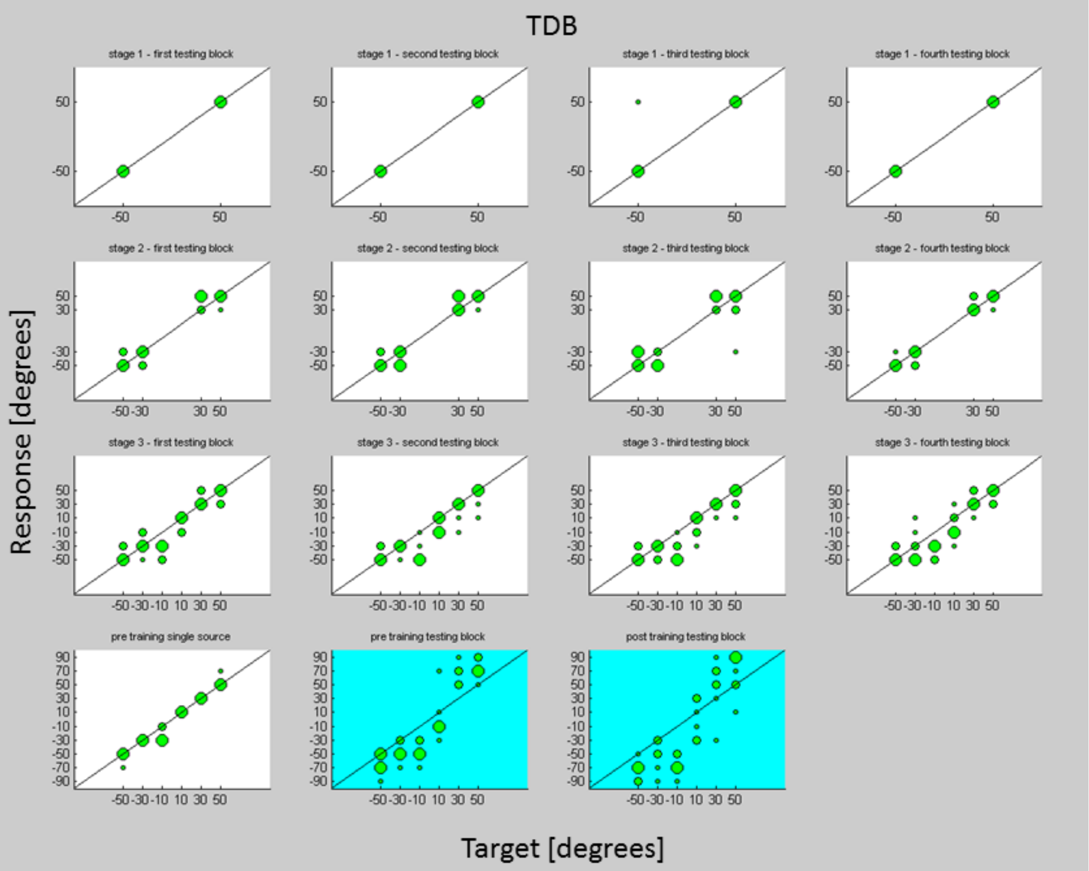

TDB
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
50
-50
-50 50
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
50 30 -30 -50
-50 -30 30 50
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
50 30 10 -10 -30 -50
-50 -30 -10 10 30 50
pre training single source
pre training testing block
post training testing block
90 70 50 30 10 -10 -30 -50 -70 -90
-50 -30 -10 10 30 50
Response [degrees]
Target [degrees]

Figure 48

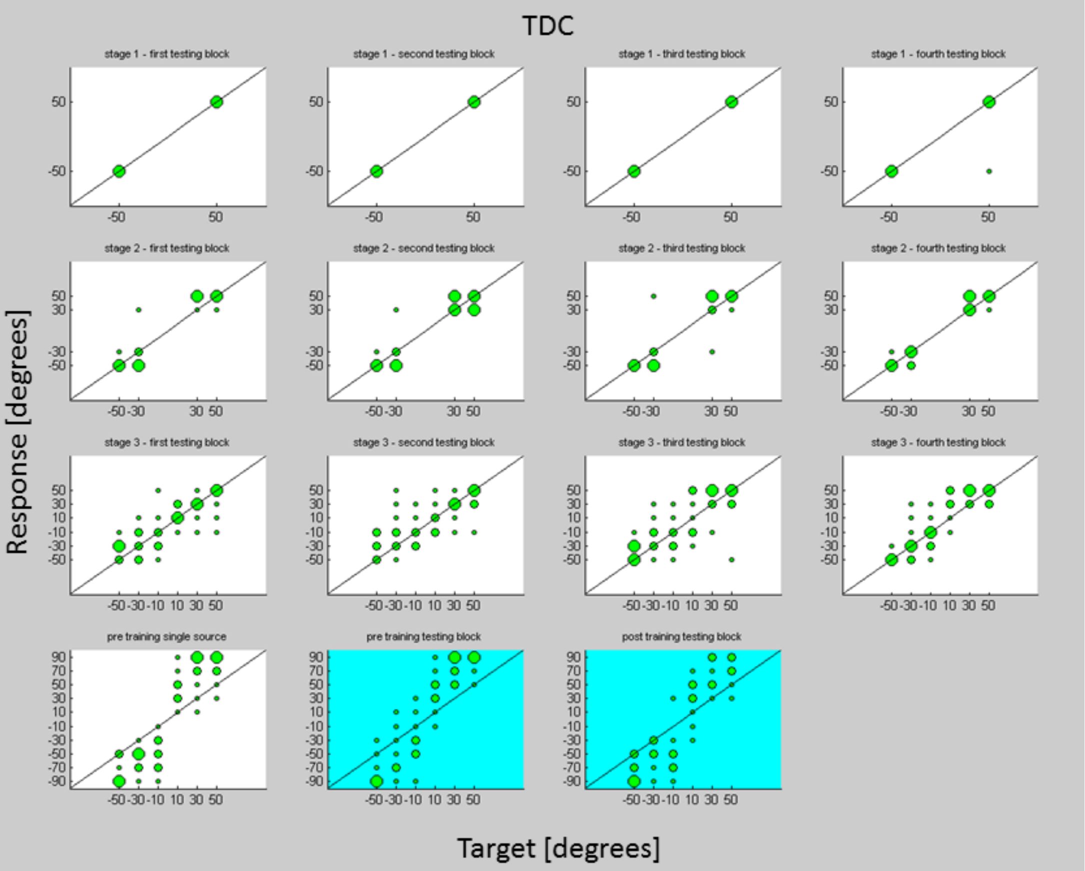

TDC
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
-90 -70 -50 -30 -10 10 30 50 70 90
-50 -30 -10 10 30 50
Response [degrees]
Target [degrees]

Figure 49

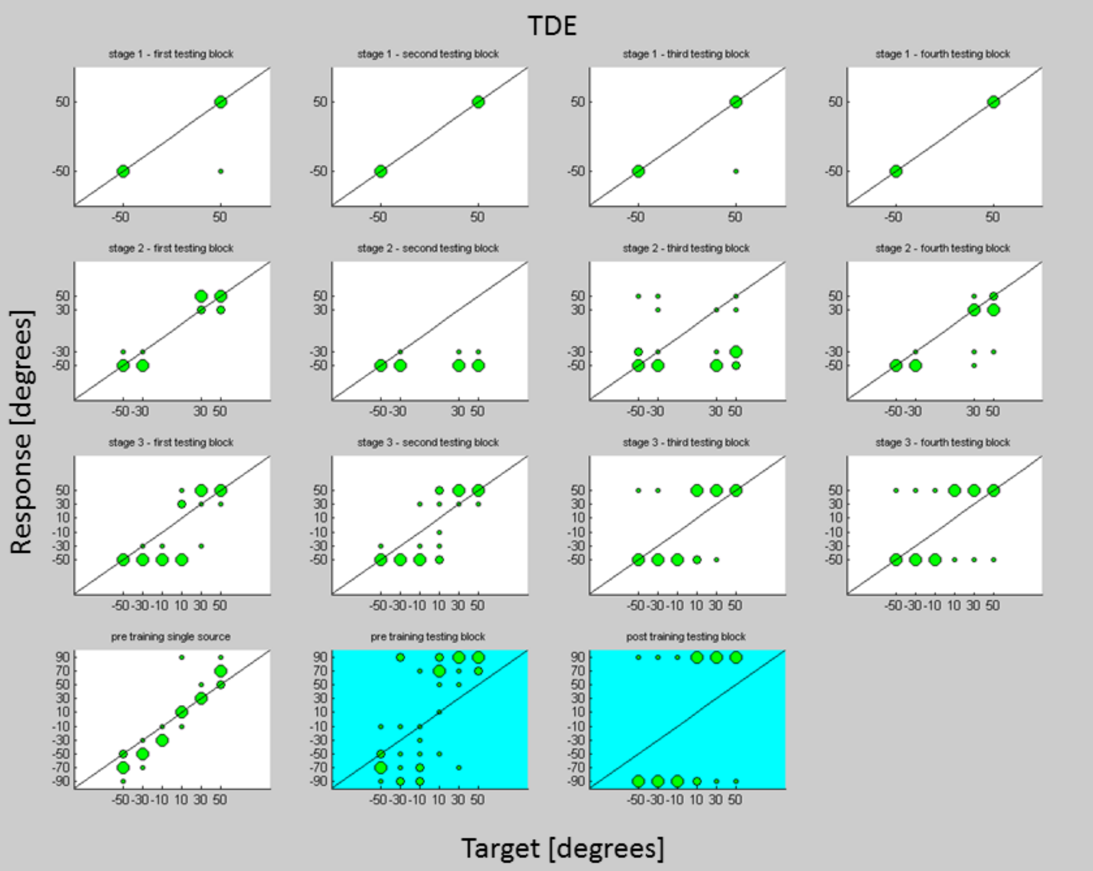

TDE
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
Response [degrees]
Target [degrees]
-50
50
-50 -30 30 50
-50 -30 -10 10 30 50
90 70 50 30 10 -10 -30 -50 -70 -90

Figure 50

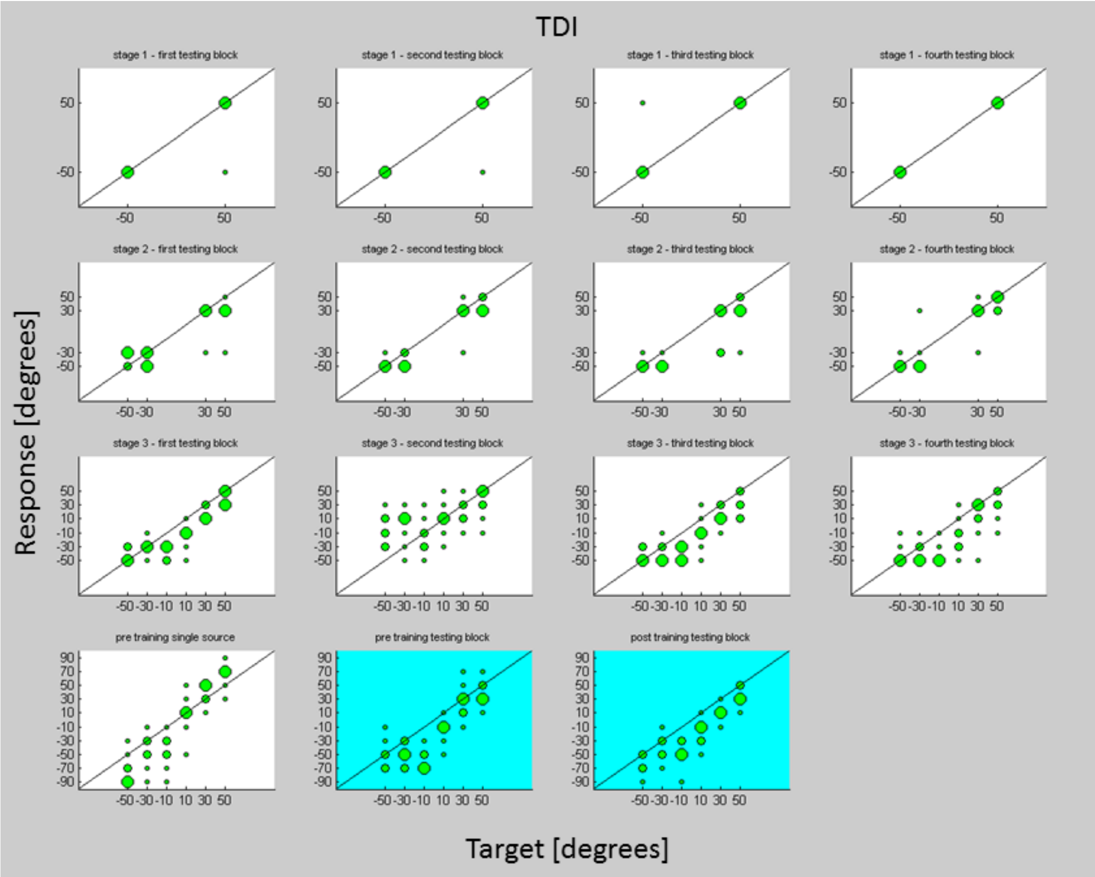

TDI
stage 1 - first testing block
stage 1 - second testing block
stage 1 - third testing block
stage 1 - fourth testing block
stage 2 - first testing block
stage 2 - second testing block
stage 2 - third testing block
stage 2 - fourth testing block
stage 3 - first testing block
stage 3 - second testing block
stage 3 - third testing block
stage 3 - fourth testing block
pre training single source
pre training testing block
post training testing block
Response [degrees]
Target [degrees]
50
-50
50
30
-30
-50
50
30
10
-10
-30
-50
90
70
50
30
10
-10
-30
-50
-70
-90
-50
50
-50 -30
30 50
-50 -30 -10 10 30 50

Figure 51

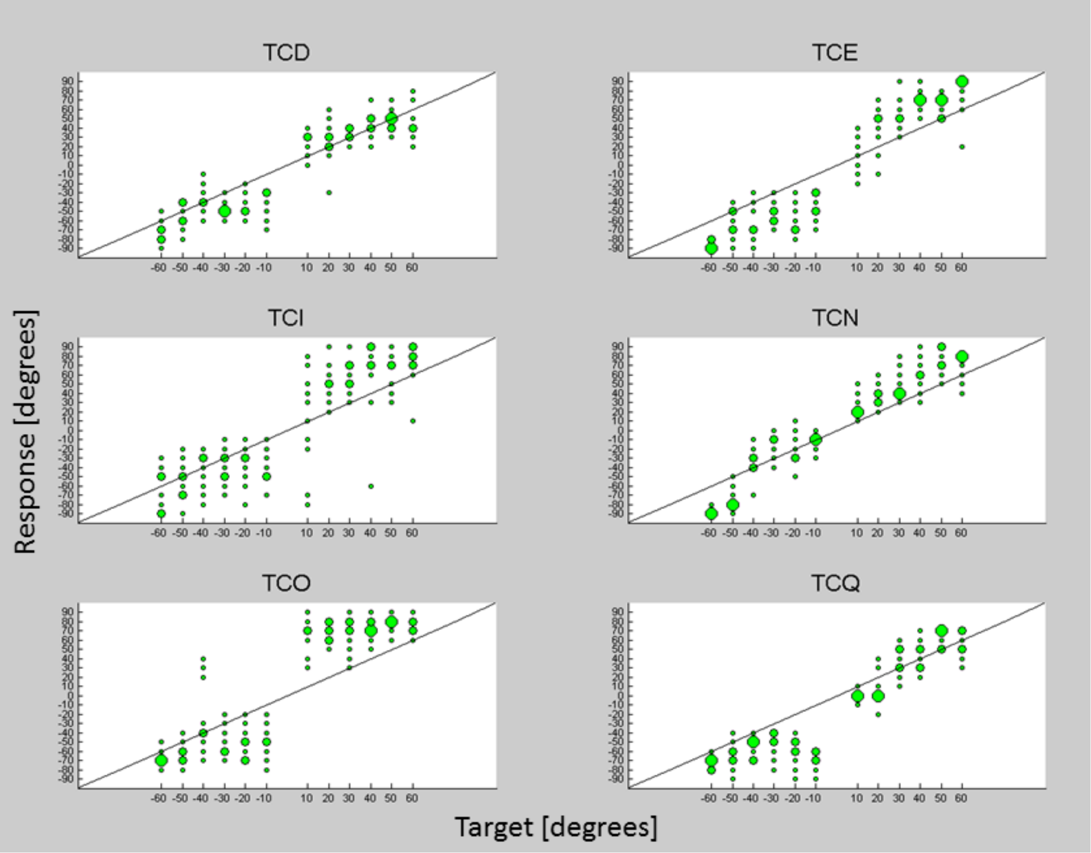

TCD
TCE
TCI
TCN
TCO
TCQ
Response [degrees]
Target [degrees]
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 52

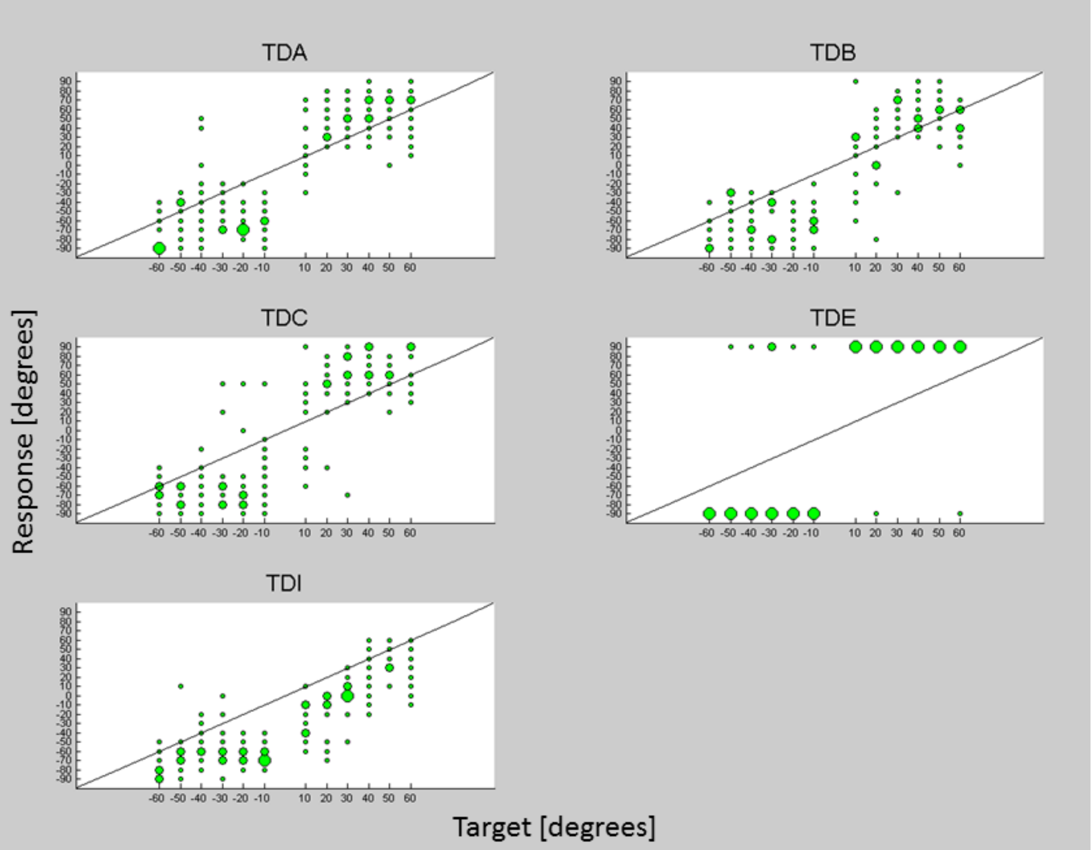


Figure 53

### Data for 'beam' stimulus

The data below describes the subjects' performance on the non-vocoded task, pre-training task, post-training task and the fine-grained task. The beam stimulus is the stimulus that the subjects were not trained on, thus there is no other "progress" data.

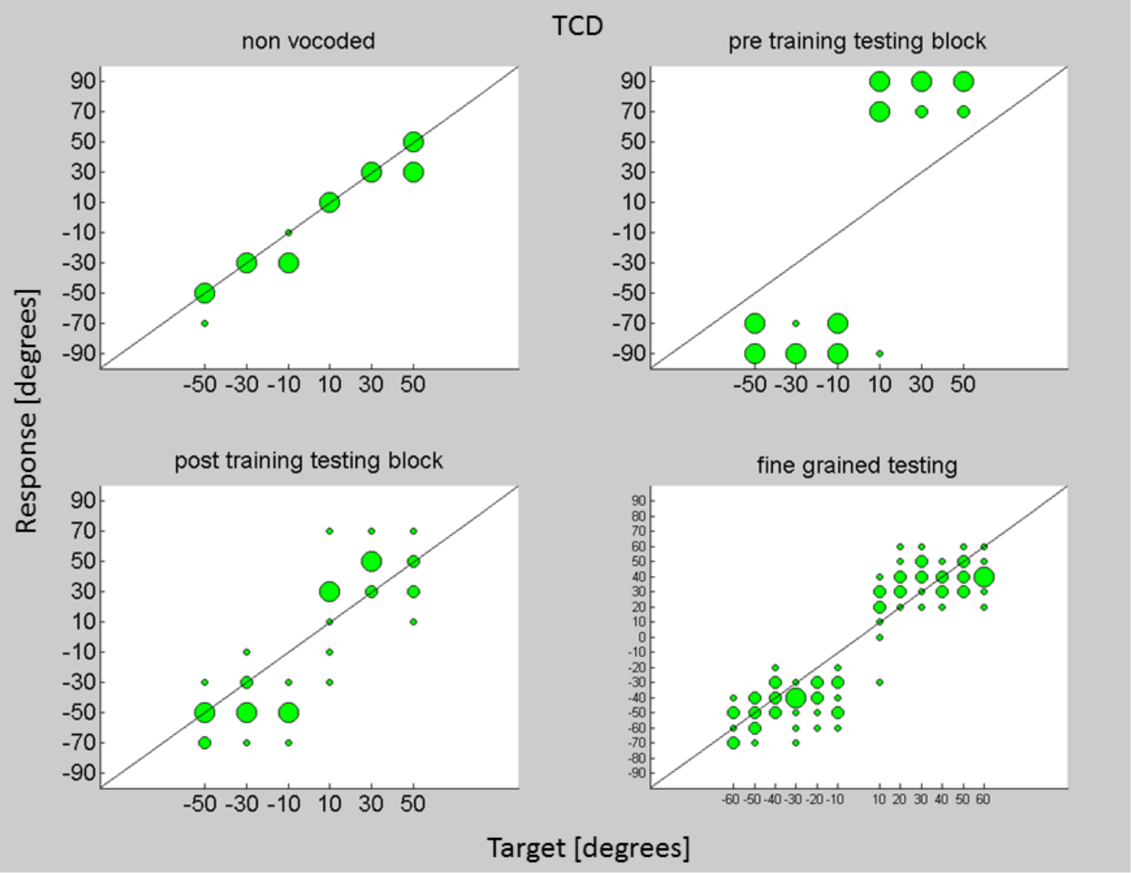

TCD
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 54

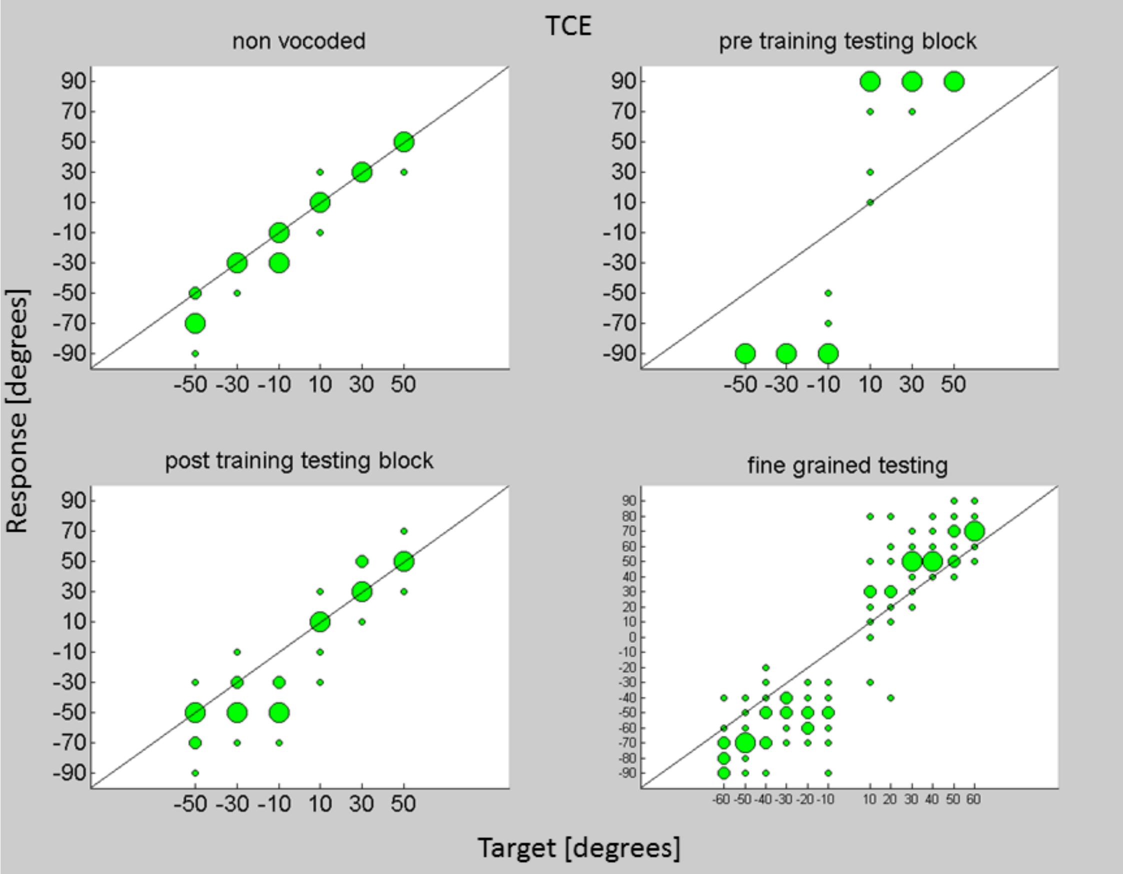

TCE
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90 70 50 30 10 -10 -30 -50 -70 -90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 55

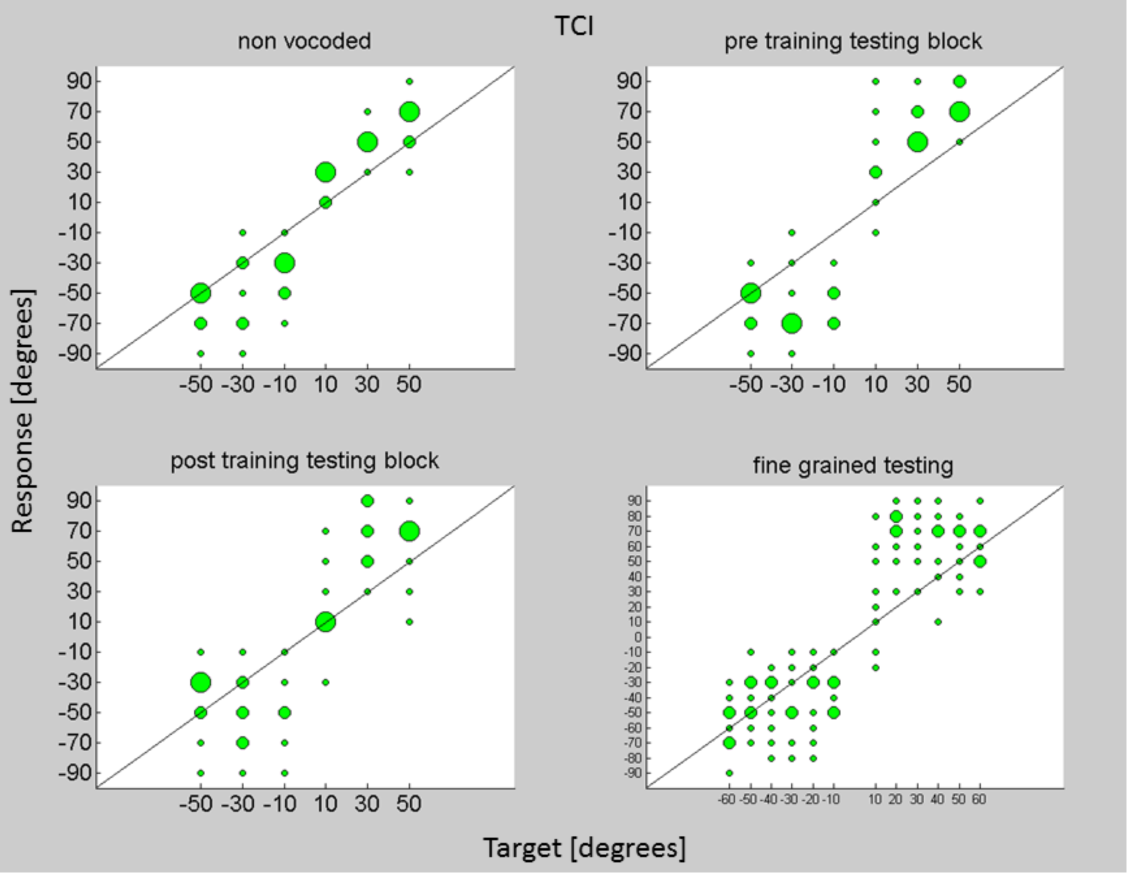

TCI
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 56

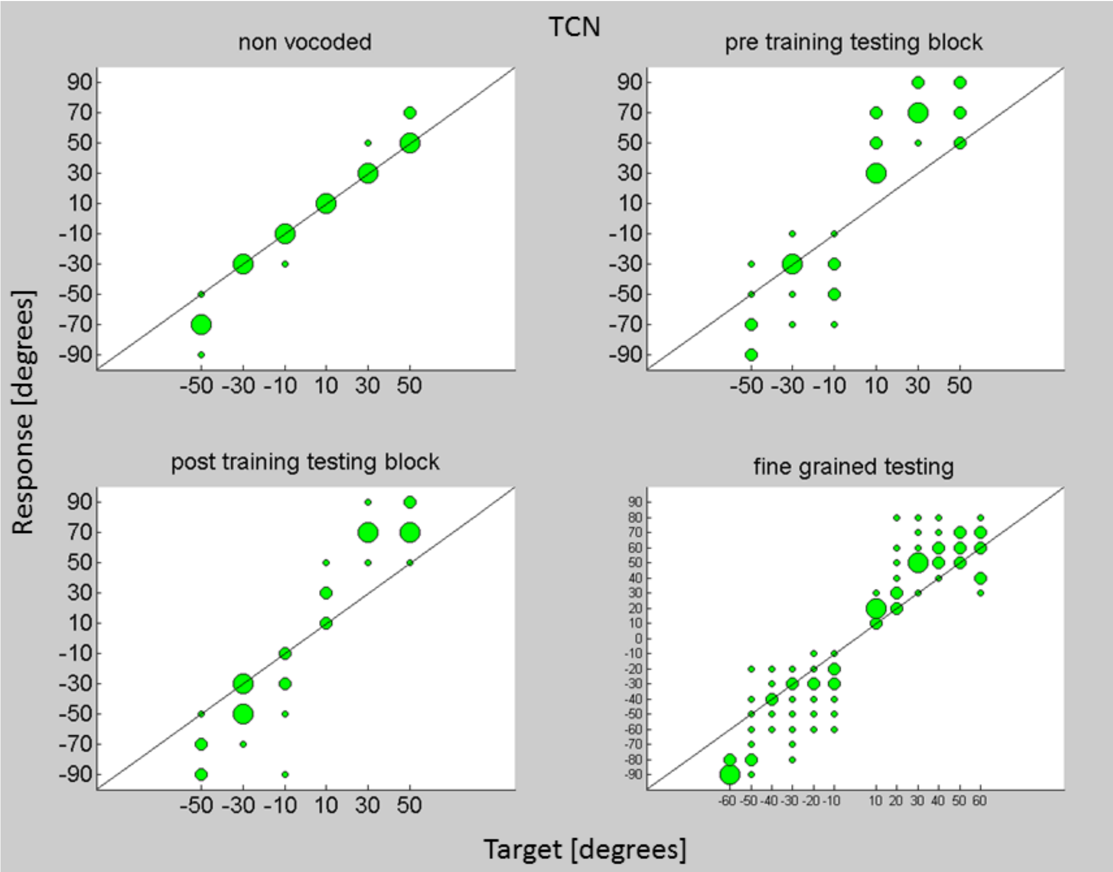

TCN
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 57

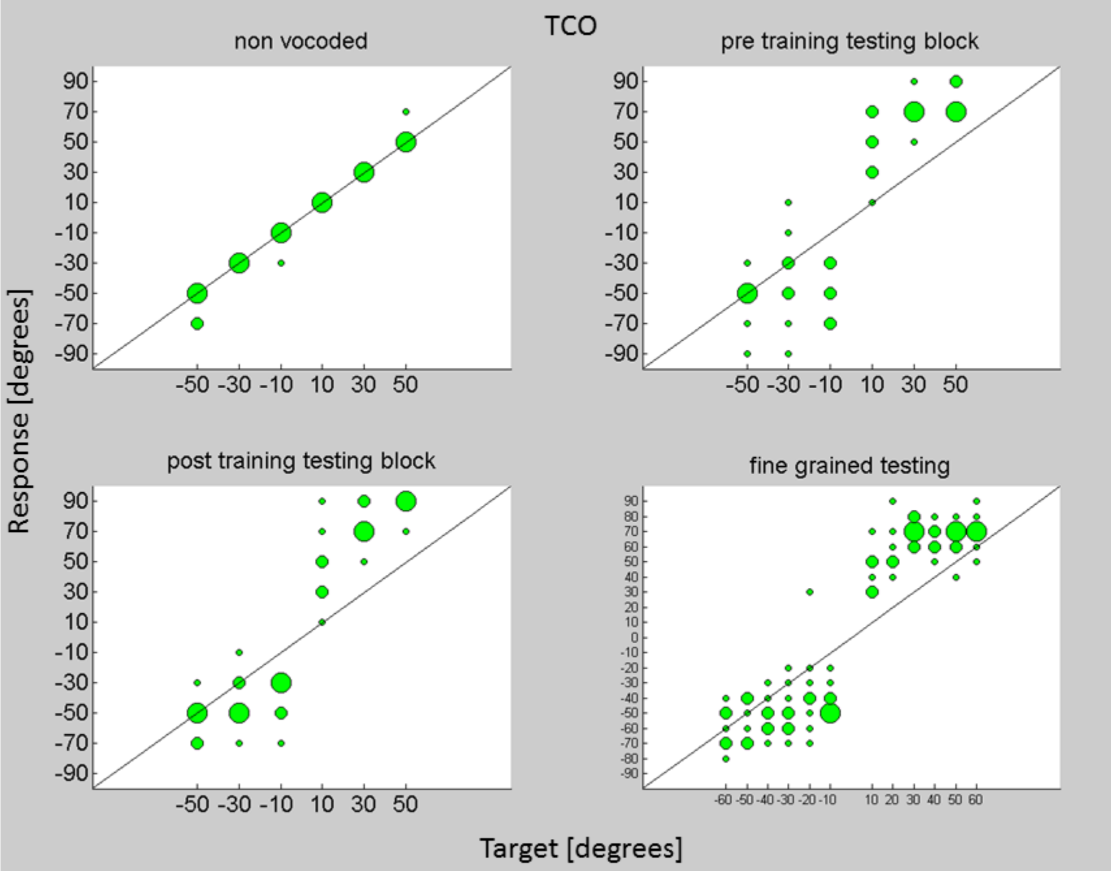

TCO
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 58

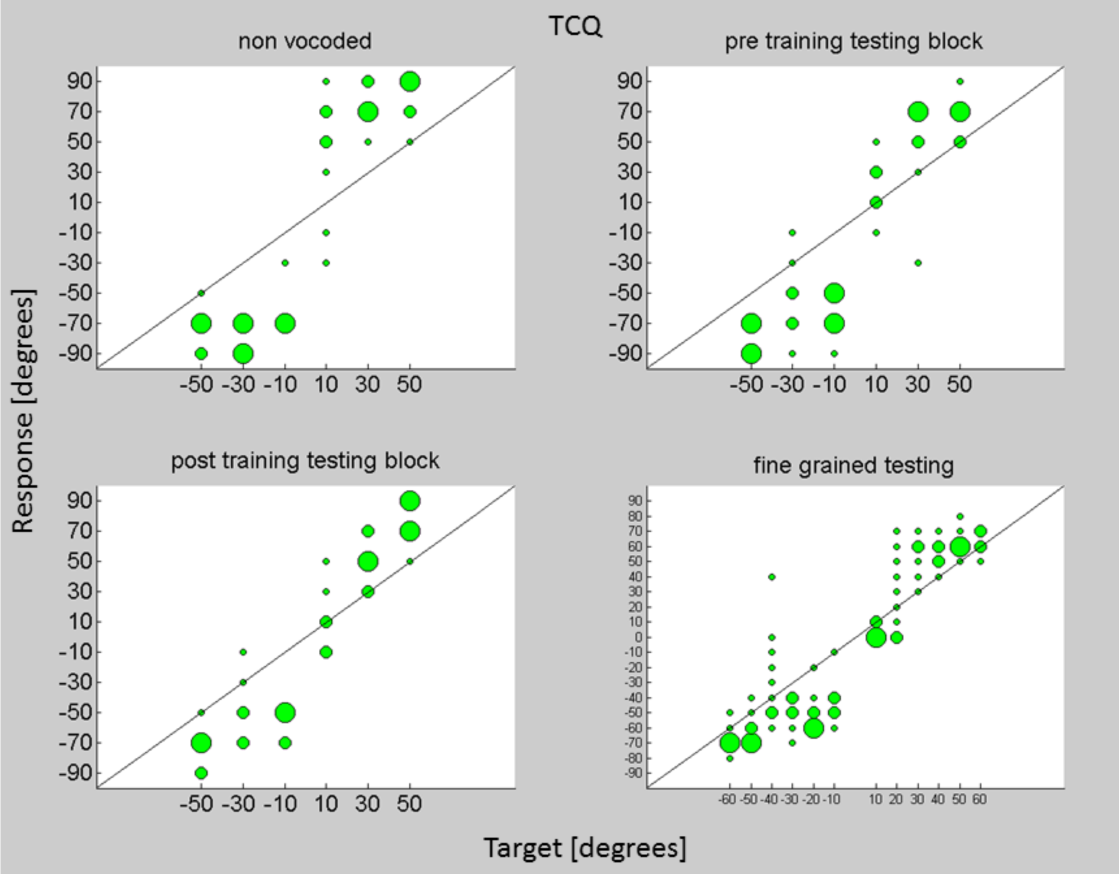

TCQ
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 59

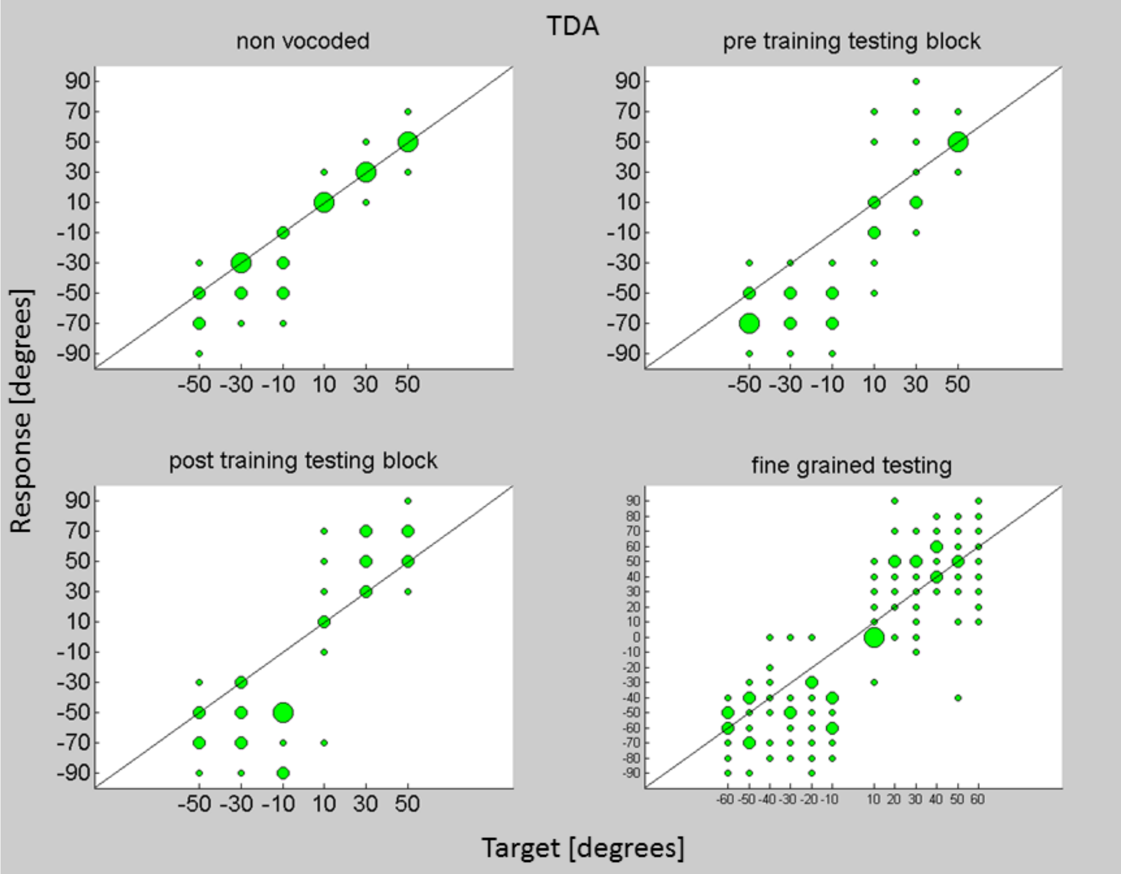

TDA
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 60

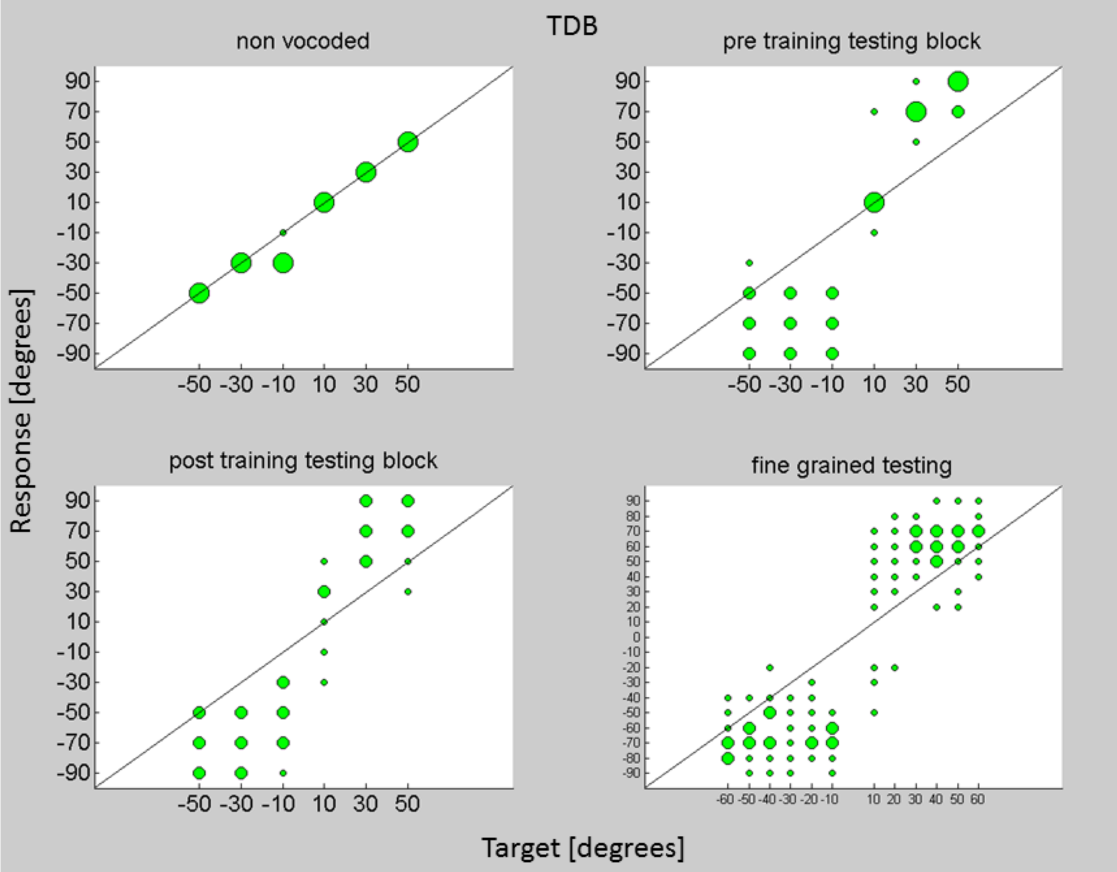

TDB
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 61

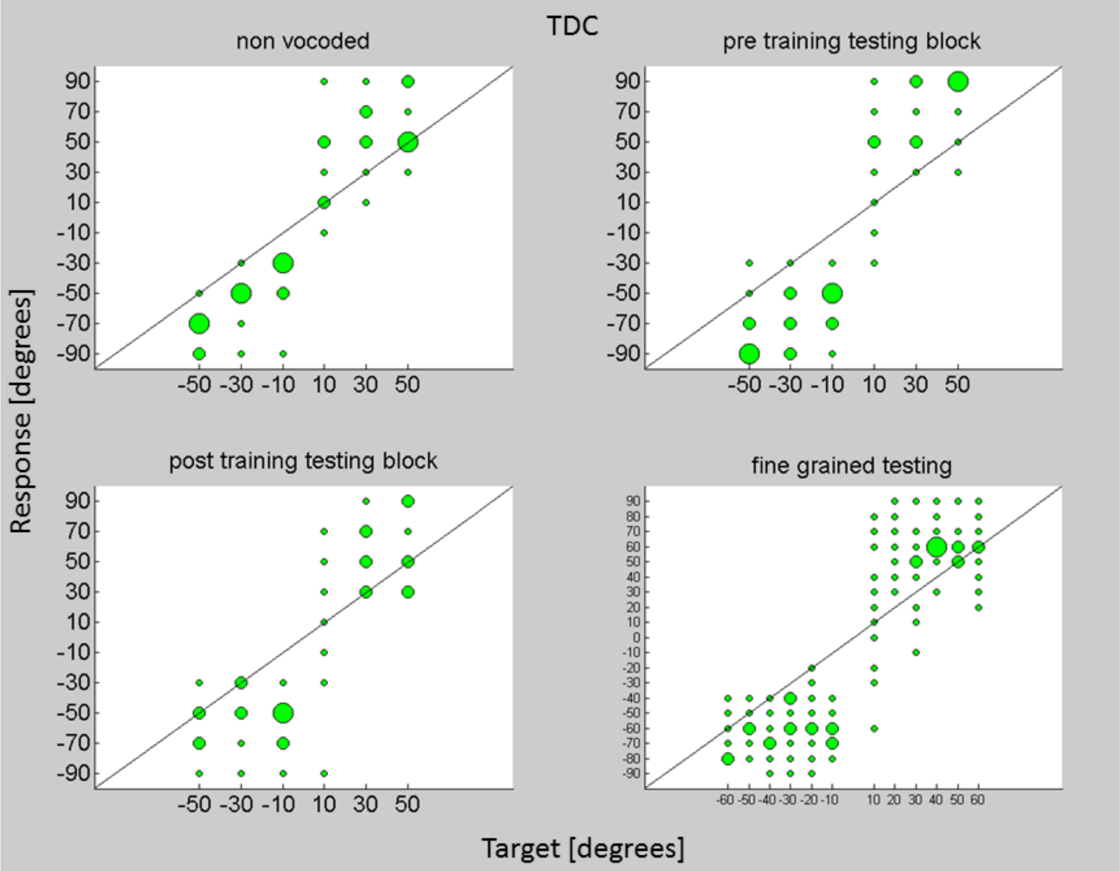

TDC
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 62

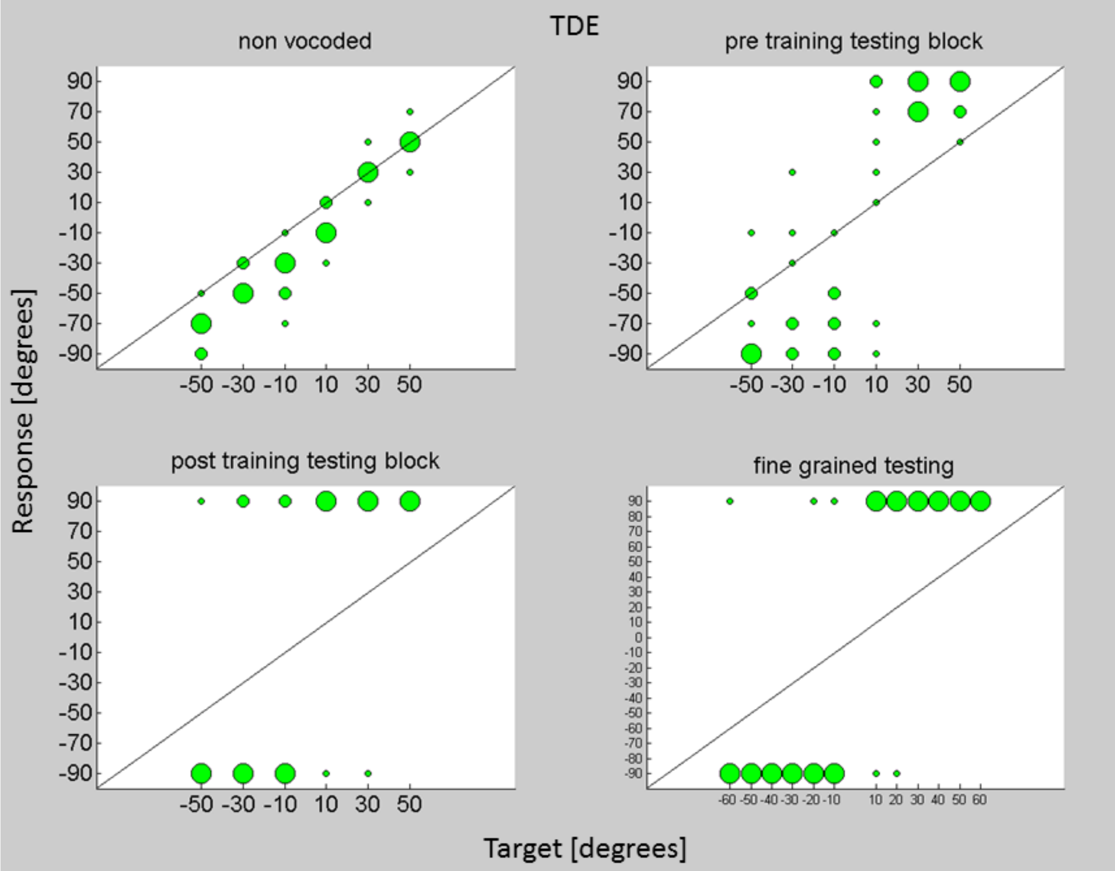

TDE
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90
70
50
30
10
-10
-30
-50
-70
-90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 63

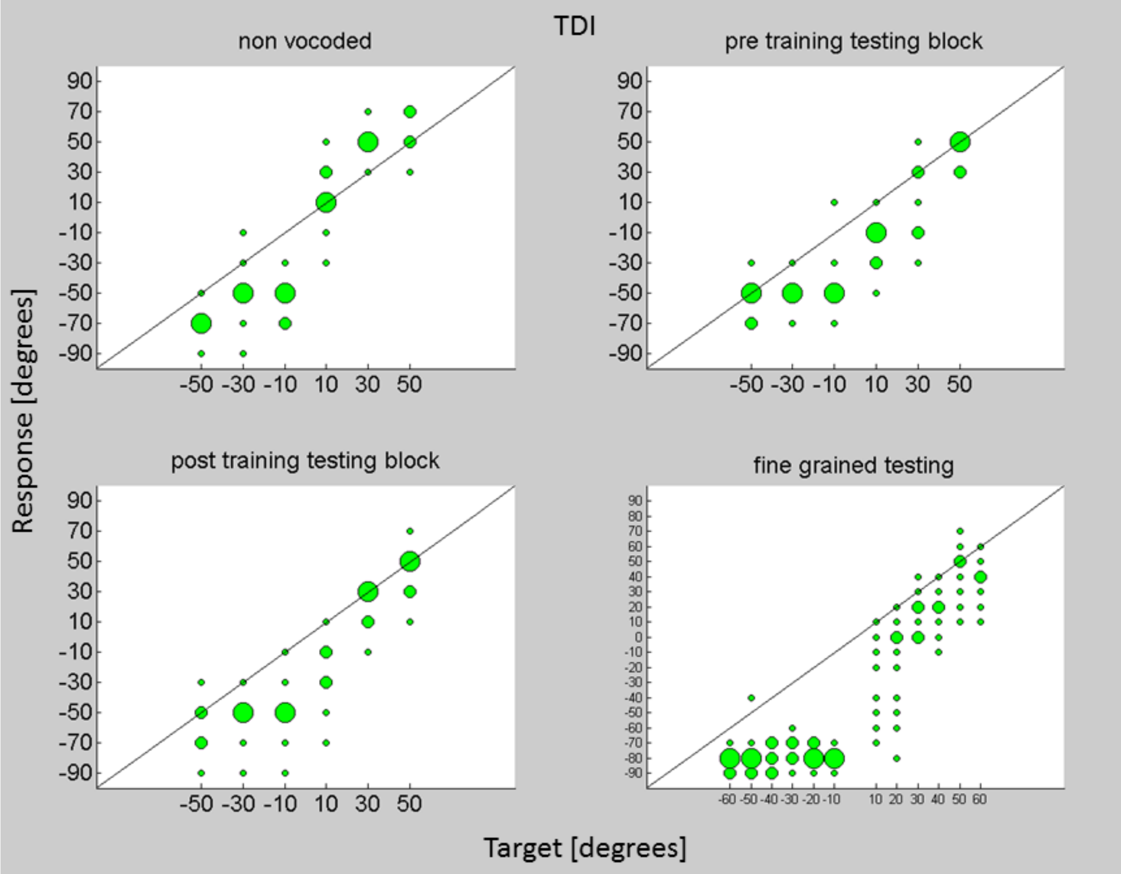

TDI
non vocoded
pre training testing block
post training testing block
fine grained testing
Response [degrees]
Target [degrees]
90 70 50 30 10 -10 -30 -50 -70 -90
-50 -30 -10 10 30 50
90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90
-60 -50 -40 -30 -20 -10 10 20 30 40 50 60

Figure 64

# BIBLIOGRAPHY

[1] R.Y. Litovsky, A. Parkinson, J. Arcaroli, Spatial hearing and speech intelligibility in bilateral cochlear implant users., Ear Hear. 30 (2009) 419–431. doi:10.1097/AUD.0b013e3181a165be.

[2] D.W. Grantham, D.H. Ashmead, T. a Ricketts, R.F. Labadie, D.S. Haynes, Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants., Ear Hear. 28 (2007) 524–541.

[3] A.C. Neuman, A. Haravon, N. Sislian, S.B. Waltzman, Sound-direction identification with bilateral cochlear implants., Ear Hear. 28 (2007) 73–82.

[4] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization by Jens Blauert, J. Acoust. Soc. Am. 77 (1985) 334. doi:10.1121/1.392109.

[5] J.C. Middlebrooks, D.M. Green, Sound localization by human listeners., Annu. Rev. Psychol. 42 (1991) 135–159. doi:10.1146/annurev.ps.42.020191.001031.

[6] R.J. van Hoesel, G.M. Clark, Psychophysical studies with two binaural cochlear implant subjects., J. Acoust. Soc. Am. 102 (1997) 495–507.

[7] R.J.M. Van Hoesel, Exploring the benefits of bilateral cochlear implants, in: Audiol. Neuro-Otology, 2004: pp. 234–246. doi:10.1159/000078393.

[8] I.M. Wiggins, B.U. Seeber, Dynamic-range compression affects the lateral position of sounds, J. Acoust. Soc. Am. 130 (2011) 3939. doi:10.1121/1.3652887.

[9] A. Kan, H. Jones, R. Litovsky, Issues in binaural hearing in bilateral cochlear implant users, in: 2013: pp. 050049–050049. doi:10.1121/1.4800192.

[10] P.C. Loizou, Mimicking the human ear, IEEE Signal Process. Mag. 15 (1998) 101–130. doi:10.1109/79.708543.

[11] W.M. Hartmann, B. Rakerd, J.B. Gaalaas, On the source-identification method, 104 (1998) 3546–3557.

[12] B.S. Wilson, M.F. Dorman, Cochlear implants: current designs and future possibilities., J. Rehabil. Res. Dev. 45 (2008) 695–730. doi:10.1682/JRRD.2007.10.0173.

[13] B.S. Wilson, M.F. Dorman, Cochlear implants: A remarkable past and a brilliant future, Hear. Res. 242 (2008) 3–21. doi:10.1016/j.heares.2008.06.005.

[14] R.J.M. van Hoesel, R.S. Tyler, Speech perception, localization, and lateralization with bilateral cochlear implants., J. Acoust. Soc. Am. 113 (2003) 1617–1630. doi:10.1121/1.1539520.

[15] F.L. Wightman, D.J. Kistler, Headphone simulation of free-field listening. II: Psychophysical validation., J. Acoust. Soc. Am. 85 (1989) 868–878. doi:10.1121/1.397558.

[16] F.L. Wightman, Headphone simulation of free-field listening . I : Stimulus synthesis, (2013) 858–867.

[17] J.M. Aronoff, D.J. Freed, L.M. Fisher, I. Pal, S.D. Soli, Cochlear implant patients' localization using interaural level differences exceeds that of untrained normal hearing listeners, J. Acoust. Soc. Am. 131 (2012) EL382.

[18] B.U. Seeber, H. Fastl, Localization cues with bilateral cochlear implants., J. Acoust. Soc. Am. 123 (2008) 1030–1042. doi:10.1121/1.2821965.

[19] B. Rakerd, W.M. Hartmann, Localization of sound in rooms, III: Onset and duration effects., J. Acoust. Soc. Am. 80 (1986) 1695–1706.

[20] C. Lorenzi, S. Gatehouse, C. Lever, Sound localization in noise in normal-hearing listeners., J. Acoust. Soc. Am. 105 (1999) 1810–1820.

[21] B. Rakerd, W.M. Hartmann, Localization of sound in rooms, II: The effects of a single reflecting surface., J. Acoust. Soc. Am. 78 (1985) 524–533.

[22] B. a. Wright, A.T. Sabin, Perceptual learning: How much daily training is enough?, Exp. Brain Res. 180 (2007) 727–736. doi:10.1007/s00221-007-0898-z.

[23] E.I. Knudsen, Instructed learning in the auditory localization pathway of the barn owl., Nature. 417 (2002) 322–328.

[24] G.H. Recanzone, Rapidly induced auditory plasticity: the ventriloquism aftereffect., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 869–875.

[25] M.P. Zwiers, a J. Van Opstal, G.D. Paige, Plasticity in human sound localization induced by compressed spatial vision., Nat. Neurosci. 6 (2003) 175–181. doi:10.1038/nn999.

[26] A.L. Giraud, E. Truy, R. Frackowiak, Imaging plasticity in cochlear implant patients, Audiol. Neuro-Otology. 6 (2001) 381–393. doi:10.1159/000046847.

[27] S.G. Lomber, M.A. Meredith, A. Kral, Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf., Nat. Neurosci. 13 (2010) 1421–1427. doi:10.1038/nn.2653.

[28] J. Rouger, S. Lagleyre, B. Fraysse, S. Deneve, O. Deguine, P. Barone, Evidence that cochlear-implanted deaf patients are better multisensory integrators., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 7295–7300.

[29] A.J. King, Visual influences on auditory spatial learning., Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364 (2009) 331–339. doi:10.1098/rstb.2008.0230.

[30] E.I. Knudsen, P.F. Knudsen, Vision calibrates sound localization in developing barn owls., J. Neurosci. 9 (1989) 3306–3313.

[31] K. Strelnikov, M. Rosito, P. Barone, Effect of audiovisual training on monaural spatial hearing in horizontal plane, PLoS One. 6 (2011) e18344. doi:10.1371/journal.pone.0018344.

[32] C.W. Bishop, S. London, L.M. Miller, Visual influences on echo suppression, Curr. Biol. 21 (2011) 221–225. doi:10.1016/j.cub.2010.12.051.Visual.

[33] R.Y. Litovsky, S.P. Godar, Difference in precedence effect between children and adults signifies development of sound localization abilities in complex listening tasks., J. Acoust. Soc. Am. 128 (2010) 1979–1991. doi:10.1121/1.3478849.

[34] R. V Shannon, F.G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, Speech recognition with primarily temporal cues., Science. 270 (1995) 303–304.

[35] R.Y. Litovsky, H.S. Colburn, W.A. Yost, S.J. Guzman, The precedence effect, J. Acoust. Soc. Am. 106 (1999) 1633–1654.

[36] H. Wallach, E.B. Newman, M.R. Rosenzweig, The precedence effect in sound localization., Am. J. Psychol. 62 (1949) 315–336. doi:10.1121/1.1917119.

[37] P.M. Zurek, The precedence effect and its possible role in the avoidance of interaural ambiguities., J. Acoust. Soc. Am. 67 (1980) 953–964. doi:10.1121/1.383974.

[38] R.L. Freyman, R.K. Clifton, R.Y. Litovsky, Dynamic processes in the precedence effect., J. Acoust. Soc. Am. 90 (1991) 874–884.

[39] X. Yang, D.W. Grantham, Cross-spectral and temporal factors in the precedence effect: discrimination suppression of the lag sound in free-field., J. Acoust. Soc. Am. 102 (1997) 2973–2983.

[40] S. Tolnai, R.Y. Litovsky, A.J. King, The precedence effect and its buildup and breakdown in ferrets and humans., J. Acoust. Soc. Am. 135 (2014) 1406. doi:10.1121/1.4864486.

[41] R.Y. Litovsky, H.S. Colburn, Precedence effects in the azimuthal and median-sagittal planes.pdf, Assoc. Res. Otol. (1997).

[42] X. Yang, D.W. Grantham, Echo suppression and discrimination suppression aspects of the precedence effect., Percept. Psychophys. 59 (1997) 1108–1117.

[43] H. Levitt, Transformed up-down methods in psychoacoustics., J. Acoust. Soc. Am. 49 (1971) Suppl 2:467+. doi:10.1121/1.1912375.

[44] T. Kawase, S. Sakamoto, Y. Hori, A. Maki, Y. Suzuki, T. Kobayashi, Bimodal audio-visual training enhances auditory adaptation process., Neuroreport. 20 (2009) 1231–1234. doi:10.1097/WNR.0b013e32832fbef8.

[45] C. McGettigan, S. Rosen, S.K. Scott, Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: implications for cochlear implantation., Front. Syst. Neurosci. 8: (2014) 18. doi:10.3389/fnsys.2014.00018.

[46] M.J. Goupell, P. Majdak, B. Laback, Median-plane sound localization as a function of the number of spectral channels using a channel vocoder., J. Acoust. Soc. Am. 127 (2010) 990–1001. doi:10.1121/1.3283014.

[47] M. Ahissar, S. Hochstein, The reverse hierarchy theory of visual perceptual learning, Trends Cogn Sci. 8 (2004) 457–464. doi:10.1016/j.tics.2004.08.011.

[48] J.F. Bergan, P. Ro, D. Ro, E.I. Knudsen, Hunting increases adaptive auditory map plasticity in adult barn owls., J. Neurosci. 25 (2005) 9816–9820. doi:10.1523/JNEUROSCI.2533-05.2005.

[49] J.S. Logan, S.E. Lively, D.B. Pisoni, Training Japanese listeners to identify English /r/ and /l/: a first report., J. Acoust. Soc. Am. 89 (1991) 874–886. doi:10.1121/1.1894649.

[50] M.I. Posner, S.W. Keele, On the genesis of abstract ideas., J. Exp. Psychol. 77 (1968) 353–363. doi:10.1037/h0028558.

[51] J.I. Gold, T. Watanabe, Perceptual learning., Curr. Biol. 20 (2010) R46–8. doi:10.1016/j.cub.2009.10.066.

[52] B.A. Dosher, Z.L. Lu, Mechanisms of perceptual learning, Neurobiol. Atten. 39 (2005) 471–476.

[53] V.R. Bejjanki, J.M. Beck, Z.-L. Lu, A. Pouget, Perceptual learning as improved probabilistic inference in early sensory areas., Nat. Neurosci. 14 (2011) 642–648. doi:10.1038/nn.2796.

[54] R. Goldstone, Perceptual Learning, Annu. Rev. Psychol. 49 (1998) 585–612.

[55] J.L. McGaugh, Memory--a century of consolidation., Science. 287 (2000) 248–251. doi:10.1126/science.287.5451.248.

[56] G. Nogaki, Q.-J. Fu, J.J. Galvin, Effect of training rate on recognition of spectrally shifted speech., Ear Hear. 28 (2007) 132–140.

[57] B. a Linkenhoker, E.I. Knudsen, Incremental training increases the plasticity of the auditory space map in adult barn owls., Nature. 419 (2002) 293–296. doi:10.1038/nature00966.1.

[58] J.J. Rieser, H.L. Pick, D.H. Ashmead, a E. Garing, Calibration of human locomotion and models of perceptual-motor organization., J. Exp. Psychol. Hum. Percept. Perform. 21 (1995) 480–497.

[59] M.J. Hacker, R. Ratcliff, A revised table of d' for M-alternative forced choice, Percept. Psychophys. 26 (1979) 168–170. doi:10.3758/BF03208311.

[60] R.K. Clifton, Breakdown of echo suppression in the precedence effect., J. Acoust. Soc. Am. 82 (1987) 1834–1835. doi:10.1121/1.395802.

[61] R.K. Clifton, R.L. Freyman, Effect of click rate and delay on breakdown of the precedence effect., Percept. Psychophys. 46 (1989) 139–145. doi:10.3758/BF03204973.

[62] D.R. Begault, 3-D Sound for Virtual Reality and Multimedia, (2000).

[63] D.D. Greenwood, A cochlear frequency-position function for several years later, (2014) 2592–2605.

[64] H. Jones, A. Kan, R.Y. Litovsky, Localization by Normal Hearing Listeners Using Individualized Head-Related Transfer Function Filtered Speech Stimuli Processed Through a Noise Vocoder, Assoc. Res. Otol. 01003083 (2012) 1003083.

[65] S. Lehmann, M.M. Murray, The role of multisensory memories in unisensory object discrimination, Cogn. Brain Res. 24 (2005) 326–334. doi:10.1016/j.cogbrainres.2005.02.005.

[66] C.T. Lovelace, B.E. Stein, M.T. Wallace, An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection, Cogn. Brain Res. 17 (2003) 447–453. doi:10.1016/S0926-6410(03)00160-5.

[67] L. Shams, R. Kim, Crossmodal influences on visual perception, Phys. Life Rev. 7 (2010) 269–284. doi:10.1016/j.plrev.2010.04.006.

[68] L. Shams, A.R. Seitz, Benefits of multisensory learning, Trends Cogn. Sci. 12 (2008) 411–417. doi:10.1016/j.tics.2008.07.006.

[69] R.S. Kim, A.R. Seitz, L. Shams, Benefits of stimulus congruency for multisensory facilitation of visual learning, PLoS One. 3 (2008). doi:10.1371/journal.pone.0001532.

[70] D. Alais, J. Cass, Multisensory perceptual learning of temporal order: Audiovisual learning transfers to vision but not audition, PLoS One. 5 (2010). doi:10.1371/journal.pone.0011283.

[71] A.R. Seitz, R. Kim, L. Shams, Sound Facilitates Visual Learning, Curr. Biol. 16 (2006) 1422–1427. doi:10.1016/j.cub.2006.05.048.

[72] A. Sand, M.E. Nilsson, Asymmetric transfer of sound localization learning between indistinguishable interaural cues, Exp. Brain Res. (2014) 1–10. doi:10.1007/s00221-014-3863-7.

[73] Y. Wang, D.M. Behne, H. Jiang, Linguistic experience and audio-visual perception of non-native fricatives., J. Acoust. Soc. Am. 124 (2008) 1716–1726. doi:10.1121/1.2956483.

[74] W. a Teder-Sälejärvi, S. a Hillyard, The gradient of spatial auditory attention in free field: an event-related potential study., Percept. Psychophys. 60 (1998) 1228–1242.

[75] W. a. Teder-Sälejärvi , S. a. Hillyard, B. Roder, H.J. Neville, Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials, Cogn. Brain Res. 8 (1999) 213–227.

[76] F. Frassinetti, F. Pavani, E. Ladavas, Acoustical vision of neglected stimuli: Interaction among spatially converging audiovisual inputs in neglect patients, J. Cogn. Neurosci. 14 (2002) 62–69.

[77] F. Leo, N. Bolognini, C. Passamonti, B.E. Stein, E. Ladavas, Cross-modal localization in hemianopia: New insights on multisensory integration, Brain. 131 (2008) 855–865. doi:10.1093/brain/awn003.