



Published in final edited form as:

Cogn Affect Behav Neurosci. 2011 March ; 11(1): 1–12. doi:10.3758/s13415-010-0006-x.

Multisensory perceptual learning reshapes both fast and slow mechanisms of crossmodal processing

Anton L. Beer,

Department of Psychology, Boston University, Boston, MA 02215, USA

Melissa A. Batson, and

Department of Psychology, Boston University, Boston, MA 02215, USA, Program in Neuroscience, Boston University, Boston, MA 02215, USA

Takeo Watanabe

Department of Psychology, Boston University, Boston, MA 02215, USA, Program in Neuroscience, Boston University, Boston, MA 02215, USA

Anton L. Beer: anton.beer@psychologie.uni-regensburg.de

Abstract

Previous research has shown that sounds facilitate perception of visual patterns appearing immediately after the sound but impair perception of patterns appearing after some delay. Here we examined the spatial gradient of the fast crossmodal facilitation effect and the slow inhibition effect in order to test whether they reflect separate mechanisms. We found that crossmodal facilitation is only observed at visual field locations overlapping with the sound, whereas crossmodal inhibition affects the whole hemifield. Furthermore, we tested whether multisensory perceptual learning with misaligned audio-visual stimuli reshapes crossmodal facilitation and inhibition. We found that training shifts crossmodal facilitation towards the trained location without changing its range. By contrast, training narrows the range of inhibition without shifting its position. Our results suggest that crossmodal facilitation and inhibition reflect separate mechanisms that can both be reshaped by multisensory experience even in adult humans. Multisensory links seem to be more plastic than previously thought.

Keywords

Perceptual learning; Multisensory; Crossmodal plasticity; Development; Visual perception; Attention; IOR

Introduction

Most species (including humans) perceive the world through several senses, such as audition or vision. Adaptive behavior requires that signals from different senses be integrated. For instance, safely crossing a street is best accomplished when our auditory percept (e.g., the roaring engine of a car) complements our visual percept (e.g., lights of the head lamps). Research over the last two decades has elucidated our knowledge of the principles that guide multisensory interactions. It is now well known that sounds modulate visual perception (Beer & Röder, 2004b, 2005; Frassinetti, Bolognini, & Ladavas, 2002; Mazza, Turatto,

Rossi, & Umiltà, 2007; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Meyer, Wuerger, Rohrbein, & Zetzsche, 2005; Spence & Driver, 1997) already at early visual processing stages (Beer & Watanabe, 2009; Eimer, 2001; McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2005). A peripheral sound cue presented briefly before a visual target (<300 ms) facilitates visual perception and enhances early brain potentials for visual stimuli appearing at the same side as the sound (Eimer, 2001; Mazza et al., 2007; McDonald et al., 2000, 2005; Spence & Driver, 1997). However, when the visual target follows the sound after a longer stimulus onset asynchrony (SOA) of more than 300 ms, visual perception is impaired (Spence & Driver, 1998a, 1998b). Despite of extensive research, it is still unclear how the fast crossmodal facilitation and the slow inhibition effect are related. The effects of sound cues on visual perception closely resemble those of visual cues. A peripheral visual cue facilitates visual perception of targets presented briefly after the cue, but impairs visual perception for longer SOAs (Posner, Rafal, Choate, & Vaughan, 1985). Originally, the fast facilitation and the slow inhibition (also known as 'inhibition of return', IOR) following a cue were thought to reflect two parts of a single biphasic attention process (Klein, 2000; Maylor, 1985; Posner et al., 1985). More recent findings suggested that the fast facilitation—at least in part—reflects energy summation of the cue and target (Tassinari, Aglioti, Chelazzi, Peru, & Berlucchi, 1994; Tassinari & Berlucchi, 1995), whereas the inhibition process reflects higher-level processing such as suppressed oculomotor responses (Taylor & Klein, 2000). For instance, Collie, Maruff, Yucel, Danckert, & Currie (2000) showed that the slow inhibition process could be temporally de-coupled from the fast facilitation process. Furthermore, brain-imaging studies showed that the cue-induced inhibition process involves brain structures known to be involved in eye movement control (Lepsien & Pollmann, 2002; Mayer, Seidenberg, Dorflinger, & Rao, 2004). It has never been tested whether crossmodal facilitation and inhibition initiated by sound cues also reflect two separate mechanisms. Given the similarities with the unimodal effects, it is likely that the fast crossmodal facilitation and the slow crossmodal inhibition reflect separate mechanisms of multisensory interactions that operate at different levels of processing.

Traditionally, the audio-visual links responsible for crossmodal interactions were thought to be innate (Bower, 1974; Gibson, 1969; Marks, 1978) or to develop early in life (Bavelier & Neville, 2002; Brainard & Knudsen, 1998; King & Moore, 1991; Knudsen, 2002; Lee et al., 2001; Wallace & Stein, 2007; Wilmington, Gray, & Jahrsdoerfer, 1994). For instance, the tectum is well known for its role in multisensory integration (Meredith & Stein, 1986a). Reorganization of audio-visual maps in the superior colliculus of the cats (Wallace & Stein, 2007) or the tectum of barn owls (Knudsen, 2002) was observed when the animals were raised in an environment of mis-aligned multisensory stimuli. However, experience of mis-aligned multisensory stimuli in adult animals failed to produce a similar reorganization of these crossmodal maps (Brainard & Knudsen, 1998; Knudsen, 2002). In humans, brain-imaging studies have shown that blind humans, when compared with sighted, showed enhanced activity in visual cortex during sound localization (Gougoux, Zatorre, Lassonde, Voss, & Lepore, 2005; Weeks et al., 2000) implying that crossmodal interactions can be altered due to experience. Some recent reports suggest multisensory plasticity even in adults. For instance, both congenitally and late blind humans showed sharper spatial tuning than sighted in peripheral sound localization (Fieger, Röder, Teder-Sälejärvi, Hillyard, & Neville, 2006; Röder et al., 1999). Adult humans consistently mis-localized sounds after wearing compressing lenses (Zwiers et al., 2003) or after experiencing sounds that were succeeded by mis-aligned lights (Shinn-Cunningham, 2000). It was proposed that crossmodal plasticity in adults reflects relatively late stages of information processing (Shinn-Cunningham, 2000). For instance, enhanced sound localization of congenitally blind humans was mediated by early brain potentials (Röder et al., 1999) whereas in late blind humans it seemed to be mediated by relatively late brain potentials (Fieger et al., 2006).

Most of these previous studies examined multisensory plasticity due to passive experience after sensory deprivation (e.g., blindness) rather than training specific aspects of multisensory interactions. As outlined above crossmodal interactions between auditory and visual perception likely reflect several partially distinct processes that operate at different levels. Previous studies examining experience-dependent plasticity did not distinguish between these processes. Therefore, the present study had two goals: First, we tested whether crossmodal facilitation and inhibition reflect separate mechanisms of multisensory processing. Second, we examined whether and how cross-modal facilitation and inhibition are modified by experience in adult humans.

Accordingly, we first examined audio-visual interactions by a crossmodal cueing task. A visual target stimulus was preceded by a sound cue with a short, medium, or long SOA. Based on previous findings, we expected crossmodal facilitation for short SOAs and crossmodal inhibition for long SOAs. As recent research on visual perception indicated that cue-induced facilitation and inhibition reflect separate processes, we examined the spatial gradient of crossmodal interactions by testing Gabor discrimination at several visual field locations (see Fig. 1a) relative to the sound source. Assuming that crossmodal facilitation and inhibition reflect separate processes, we expected that sounds facilitate visual perception only at visual field locations that overlap with the sound source. By contrast, the slower inhibition effect, which is thought to be mediated by suppressed oculomotor responses (Collie et al., 2000; Lepsien & Pollmann, 2002; Mayer et al., 2004; Tassinari & Berlucchi, 1995; Taylor & Klein, 2000), should be observed across the whole hemifield.

Subsequently, we tested whether multisensory experience modifies these crossmodal interactions. Therefore, observers underwent eight perceptual learning sessions before they were re-tested with the crossmodal cueing task (Fig. 1b). During multisensory perceptual learning sessions, we aimed to establish a new spatial link between sounds and visual targets (Fig. 1c). More specifically, we were curious whether it is possible to reshape multisensory interactions in a way that sounds that originally affected perception at aligned visual field locations would affect perception at other visual field locations. Accordingly, the training sessions established a contingency between the sounds and a mis-aligned aperture location. We adopted a so-called task-irrelevant perceptual learning paradigm (Watanabe & Sasaki, 2001), which is known to primarily affect low-level visual processing (Watanabe et al., 2002). If multisensory interactions can be reshaped by experience, we would expect that after training sounds affect visual perception at the trained aperture location. In addition, it is possible that the effects of sounds on aperture locations that were generically aligned with the sound disappear.

Methods

Participants

Twenty-five paid (\$8 per session) students (normal hearing and vision) volunteered after giving written informed consent. Seven participants quit before finishing all sessions. All of the remaining 18 participants (age 20 to 33, six male) were right-handed. The study was approved by the Institutional Review Board of Boston University.

Apparatus

Participants fixated a bull's eye at the center of a CRT monitor (40×30 cm, 1280×1024 pixels, 75 Hz) in a dark room. Their head was stabilized by a chin rest and a nose clamp at a viewing distance of 60 cm. Two small speakers were mounted to the left and right vertically aligned with fixation (Fig. 1a). Stimuli were presented with Psychophysics Toolbox

(Brainard, 1997) version 2.55 and MATLAB 5.2.1 (MathWorks, Natick, MA) on a G4 Macintosh computer (OS 9).

Stimuli

Visual stimuli were oriented (45° or 135°) Gabor patches (diameter 6° visual angle, 200 ms). A sinusoidal grating (maximum luminance 11.76 cd/m^2 , spatial frequency 1.0 cycle/°) degraded by noise (60% of pixels randomly displaced) was faded to the black background (0.01 cd/m^2) by a two-dimensional Gaussian (SD of 1.5°). The noise ratio was determined in pilot studies and was chosen to yield performance rates in the Gabor discrimination task that were slightly above the psychophysical threshold for most observers. Gabors were presented at 16° visual angle eccentricity either on the left or right at one of five vertical locations (Fig. 1a) that were either aligned (A), proximal by 6° (P) or distal by 12° (D) to the sound source.

Auditory stimuli were white noise sounds presented with a sound pressure level of about 80 dB (A) as measured at ear position. Free field sounds (rather than headphone sounds) were used as they best preserve the spectral cues necessary for vertical localization (Middlebrooks & Green, 1991). Moreover, previous research suggested that low-level multisensory interactions likely require both binaural and monaural localization cues (Alais & Burr, 2004; Meyer & Wuerger, 2001; Meyer, Wuerger, Rohrbein, & Zetzsche, 2005). The sounds were presented through the two speakers that were vertically aligned with fixation. However, due to the monitor chassis the two speakers were slightly lateral to the screen position of the central aperture location (A). Close spatial overlap between auditory and visual stimuli is crucial for low-level crossmodal mechanisms (e.g., Meredith & Stein, 1986b; Meyer et al., 2005; Frassinetti et al., 2002). Therefore, we corrected the azimuth locations of the sounds by adjusting the interaural level differences (ILDs) according to the law of sines (Bauer, 1961). By means of this adjustment, a single sound was perceived at an azimuth location overlapping with the mid-vertical screen apertures (A) at the left or right side, respectively. Adjusting the ILDs by the law of sines has been successfully used in previous auditory localization research (e.g., Beer & Röder, 2004a, 2004b; Grantham, 1986). Note that this procedure preserves spectral cues and corrects for ILDs, but does not correct for interaural time differences (ITDs). However, we considered the error on ITDs negligible for several reasons: Sounds only had to be adjusted by about 5° of azimuth angle. Humans only use ITDs on low-frequency components (less than about 1 to 1.5 kHz) of the sounds (Middlebrooks & Green, 1991). We used white noise sounds with a wider spectrum of higher frequencies ($>1.5 \text{ kHz}$) than lower frequencies. We were primarily interested in vertical shifts of crossmodal interactions. Vertical sound localization primarily relies on spectral cues (Middlebrooks & Green, 1991).

Procedure

Audio-visual interactions were tested prior to and following eight training sessions (Fig. 1b). Each of the two test sessions and eight training sessions lasted about 1 h and was conducted on a separate day.

Tests—Test trials started with the presentation of a bull's eye as fixation point in the screen center. After a variable time interval between 400 and 650 ms, a sound was presented for 100 ms (2-ms rise and fall time) either on the left or right aligned with the mid-vertical visual aperture location (A). With a SOA of either 150 ms, 300 ms, or 1,000 ms, the sound was followed by an oriented Gabor patch that appeared for 200 ms at either the same (valid) or opposite (invalid) side. Previous research has shown that the slow inhibition effect following a sound cue may depend on a central auditory reorienting event between the sound cue and the target stimulus (Spence & Driver, 1998a, 1998b). This auditory

reorienting event is comparable to a fixation spot in visual cueing paradigms. As peripheral cues may attract attention and eye movements to the cued location and therefore mask possible inhibition effects, the fixation spot or central auditory reorienting event are meant to unmask inhibition effects by refocusing attention and eye movements towards the center. Therefore, on trials with a long SOA (1,000 ms) an additional central white noise sound (50 ms, 2-ms rise and fall time) was presented 300 ms after the onset of the sound cue. As we were interested in the spatial gradient of crossmodal interactions, the Gabor either completely overlapped with the sound source or appeared above or below. In particular, the Gabor was presented at one of five vertical locations relative to the sound source: one aligned (A) with, and two each (one above and one below the horizon) proximal (P) or distal (D) to the sound source. Each sound-Gabor pair was equally likely. Observers were asked to judge the orientation of the Gabors. This was done in order to avoid response biases (Spence & Driver, 1997) that may occur when observers have to judge the location of visual targets (Shinn-Cunningham, 2000). Observers responded by pressing one of two keys corresponding to the two possible orientations (45° or 135°). They were asked to respond both fast and accurate. In order to deal with outliers only responses occurring within 250 ms and 900 ms after Gabor onset were analyzed. The inter-trial-interval varied randomly between 1,250 ms and 1,500 ms. The test session consisted of four blocks of 240 trials. No feedback was provided during or after these blocks. A brief practice block at the beginning of each session served to familiarize participants with the task. Eye movements were recorded on a sub-sample of participants (see below). For the analysis, six datasets had to be excluded due to chance performance ($< 75\%$ correct) or extensive eye movements ($> 10\%$ trials not fixated) at the first test session. Note that the primary dependent measure was discrimination performance and that datasets with chance (or ceiling) performance cannot be interpreted.

Training—Our multisensory training consisted of a so-called task-irrelevant perceptual learning paradigm (Watanabe et al., 2001, 2002; Seitz & Watanabe, 2003). With this paradigm, learning can be directed towards a specific stimulus configuration while the exposure rate can be kept constant. This is particularly important when unimodal and multisensory learning effects need to be separated. For instance, if training consisted of only one sound-Gabor pair, changes in the perception for the exposed Gabor location could be due to purely visual perceptual learning rather than altered audio-visual links. With task-irrelevant perceptual learning, the task-irrelevant stimulus is presented together with another easy to detect target stimulus. Observers are asked to detect the target stimulus while ignoring the learning stimulus. It has been shown that stimuli that were task-irrelevant during training but that were paired with a relevant target were perceived more accurately after training than before (Seitz & Watanabe, 2003).

During training sessions, participants perceived a rapid stream of sounds each lasting 100 ms (2-ms rise and fall time) with an inter-sound-interval varying randomly between 550 and 750 ms. Sounds appeared either on the left or right aligned with the mid-vertical visual aperture location (A). Together with each sound (no delay) a Gabor was presented for 200 ms at either a proximal or aligned aperture location. All sound-Gabor pairs were equally likely. After a delay of 150 ms relative to the onset of the sound-Gabor pair a circle or square encompassed the Gabor for 50 ms. Observers were asked to fixate the center of the display and detect an infrequent (25%) target shape (circle or square, alternating across sessions) by pressing a button while ignoring the sounds or Gabors. They were encouraged to respond fast. Responses occurring within 100 ms and 700 ms after shape onset were assigned to the preceding shape as hits or false alarms, respectively. Target shapes appeared equally often at all locations. However, target shapes established a contingency between same-side sounds and Gabors at the trained (P_l) locations at either side (Fig. 1c) while targets preceded by Gabors at other locations (A/P_u) were equally often paired with left or

right sounds. Note that this protocol presents all sounds equally often on both sides (left or right A) and all visual stimuli equally often at all aperture locations (left or right $P_u/A/P_l$). Therefore, training induced changes in visual discrimination may not be attributed to unimodal stimulus probabilities. Instead, the contingency between sounds and Gabors at the trained location was solely established by the relevance of the simple shape. Training consisted of eight sessions that were conducted on separate days. Each session consisted of six blocks of 576 trials each. Performance feedback was provided after each block. For half of the participants, the to-be-trained proximal aperture locations (P_l) were either above or below (left and right side, respectively) the aligned locations. Mirror-symmetric aperture locations (lower on the left and upper on the right side, respectively) were trained on the remaining participants. Eye movements were recorded on a sub-sample of participants (see below). One dataset had to be excluded from the analysis, because of chance performance ($d' < 1.35$) at the first training session.

Eye movement recording—Eye movements were continuously recorded (calibration window 90% of screen, 30-Hz digitization rate) from the right eye of nine participants with a View Point QuickClamp Camera System (Arrington Research, Scottsdale, AZ). Eye movement recording was omitted on the other observers because of technical reasons or difficulties in adequately calibrating the eye tracker. We derived several measures of eye movements: fixation stability, gaze deviation towards the visual stimulus, and gaze deviation towards the sound. In order to equate for differences in trial duration (e.g., 1,000 ms SOA trials vs. 150 ms SOA trials) we only examined eye position measures during the presentation of the visual stimuli. A trial was considered not fixated when the point of gaze on any measurement throughout stimulus presentation deviated more than 5° from the fixation point. Gaze deviation towards the visual stimulus was assessed by calculating for each eye movement measurement the deviation of the point of gaze from fixation towards the side of the visual stimulus (e.g., Gabor or shape). The largest deviation was taken as the best estimate for gaze deviation. Similarly, gaze deviation towards the sound was assessed by calculating on each trial the largest deviation of the point of gaze from fixation towards the sound side. No temporal filtering was applied.

As we were only able to record eye movements from nine of the participants, there is a risk that the results differ for observers with and without eye movement recordings. Therefore, we repeated all analyses reported in the results section including the factor Eye Movement Recording (conducted vs. omitted). We observed no significant differences between the data from participants with and without eye movement recordings. Therefore, datasets were pooled.

Results

Test prior to training

Audio-visual interactions were tested prior to training for five vertical aperture locations (A , P_l , P_u , D_l , D_u) and three SOAs (150 ms, 300 ms, 1,000 ms). The sound cue appeared either on the same (valid) or opposite (invalid) side as the Gabor. We first examined the crossmodal validity effect (difference in discrimination performance for validly versus invalidly cued Gabors) for the first test session (prior to training). A within-subjects ANOVA on percentage correct (Fig. 2a) revealed significant interactions of $SOA \times Validity$, $F(2, 22) = 3.6$, $p = .043$, and $SOA \times Vertical Location \times Validity$, $F(8, 88) = 2.6$, $p = .015$. In order to resolve these interactions, separate analyses were conducted for each SOA level.

Gabors following sounds after 150 ms were discriminated more accurately with sounds appearing at the same (valid) side than at the opposite (invalid) side. However, a within-subjects ANOVA revealed a significant interaction of $Validity \times Vertical Location$, $F(4, 44)$

$= 2.6, p = .050$. Subsequent paired-samples t tests indicated that crossmodal facilitation emerged only for Gabors at visual field locations that overlapped with the sound source (A), $t(11) = 2.5, p = .028$, but not at closeby aperture locations (P, D).

As illustrated in Fig. 2a, the crossmodal facilitation effect observed at short SOAs at the aligned location was absent for trials with intermediate sound-Gabor delays (300 ms). A within-subjects ANOVA on Gabor discrimination performance for the 300 ms SOA revealed no significant main effect or interaction.

Gabors following the sound after 1,000 ms were less accurately discriminated on the same side (83.3%) as the sound compared to Gabors appearing at the opposite side (85.1%)—reflecting crossmodal inhibition. This was confirmed by a within-subjects ANOVA on percentage correct Gabor discriminations, which revealed a main effect of Validity, $F(1, 11) = 8.6, p = .014$. However, no main effect or interaction with the vertical location of the Gabor patches was observed, suggesting that sounds impaired visual discrimination equally across all tested aperture locations (A, P, D).

Note that Gabors were embedded in noise and the noise level was chosen to provide performance levels slightly above threshold. Therefore, we primarily expected cross-modal cueing effects on the discrimination performance. Nevertheless, in order to rule out potential speed-accuracy trade-offs we also examined response times. Responses to Gabors tended to be equally fast across all aperture locations regardless of the validity of the sound cue. A within-subjects ANOVA including the factors SOA, Vertical Location, and Validity revealed no main effect or interaction on response times, suggesting that speed-accuracy trade-offs may not account for the VE (cross-modal facilitation or inhibition, respectively) observed on discrimination performance (see above). It might be argued that poor statistical power prevented us from detecting potential speed-accuracy trade-offs. However, as can be seen in Fig. 2b differences in response times (if there were any) tended to be complementary to the effects on discrimination performance. For instance, Gabors presented at the 1,000-ms SOA were less accurately discriminated when the preceding sound appeared in the same hemifield than in the opposite hemifield. Complementary to this impaired discrimination performance, responses also tended to be slower when Gabors appeared in the same hemifield (718 ms) than in the opposite hemifield (711 ms) of the preceding sound, $F(4, 44) = 4.5, p = .056$.

There is a risk that the validity effects reported above were the result of retinal shifts due to eye movements. For instance, if observers involuntarily directed their eyes to the sound source, subsequent Gabors presented at the sound location were more discriminable simply because they appeared in the central visual field. Therefore, we analyzed eye movements by a three-way within-subjects ANOVA (SOA, Vertical Location, Validity). Observers fixated the center of the screen fairly well. On average, only 2.5% of the trials did not meet our fixation criterion. Most important, we found no significant main effect or interaction on either of the eye movement measures (proportion of not fixated trials, gaze deviation towards the Gabor or sound). This suggests that the validity effects may not be attributed to eye movement induced shifts of the Gabor patches into the central visual field.

Test differences (post-training versus pre-training)

We were interested in how multisensory perceptual learning affected audio-visual links that give rise to crossmodal facilitation or inhibition. Therefore, we compared Gabor discrimination performance after training versus the performance before training by a within-subjects ANOVA on percentage correct including the factors Test Session, SOA, Vertical Location, and Validity. This analysis revealed a significant interaction of $SOA \times Validity$, $F(2, 22) = 5.04, p = .016$, as well as a significant four-way interaction of Test

Session \times SOA \times Vertical Location \times Validity, $F(8, 88) = 2.05, p = .049$. No other main effects or interactions were significant. In order to resolve the four-way interaction, we compared the change in the validity effect across test sessions separate for each SOA level. This effect is illustrated in Fig. 3, which depicts the crossmodal VE as found in the test sessions pre- and post-training (Fig. 3a) and the change of the crossmodal VE from pre- to post-training (Fig. 3b).

For the short 150-ms SOA, training resulted in a spatial shift of crossmodal facilitation from the aligned (A) to the trained (P_t) location. A within-subjects ANOVA revealed a significant interaction between Test Session \times Vertical Location, $F(4, 44) = 1.9, p = .030$. Subsequent paired-samples t tests indicated that the VE reflecting crossmodal facilitation increased at the proximal trained (P_t) location, $t(11) = 2.4, p = .037$, and decreased at the aligned (A) location, $t(11) = 2.2, p = .050$. No changes were observed at the untrained proximal (P_u) or distal (D_u, D_t) control locations. No significant effects were observed for the intermediate 300-ms SOA.

For the long 1,000-ms SOA, crossmodal inhibition did not change from pre-training to post-training for either aligned or proximal locations. Separate t tests for the aligned (A), mis-aligned trained (P_t), or the mis-aligned untrained control (P_u) locations indicated no significant differences. However, crossmodal inhibition at post-training was significantly reduced for the distal Gabor apertures (D_u/D_t) compared to the aligned or proximal Gabor apertures ($P_u/A/P_t$), $t(11) = 3.3, p = .007$.

Again, we tested for differences in response times with a four-way within-subjects ANOVA including the factors Test Session, SOA, Vertical Location, and Validity. Observers responded on average faster in the post-training test (689 ms) than in the pre-training test (716 ms) as revealed by a main effect of Test Session, $F(1, 11) = 6.10, p = .031$. No other main effects or interactions were significant. The analysis of eye movements indicated no main effect or interaction on either number of not fixated trials, gaze deviation towards the Gabor, or gaze deviation towards the sound.

Training

During multisensory perceptual learning sessions, observers had to detect a simple target shape (e.g., square) among a stream of distractor shapes (e.g., circle). For the analysis of training performance, two consecutive sessions were pooled, because the target shape (circle or square) alternated across sessions. Note that the task-relevant stimulus (shape) was relatively salient. Accordingly, learning effects were expected on both response times and detection performance. Performance levels (percentage of hits) on simple detection tasks are susceptible to criterion shifts. Therefore, we calculated the sensitivity measure d' (Swets, 1973) in order to obtain a criterion neutral measure of shape perception. Shapes could appear at three different apertures ($P_u/A/P_t$). At two apertures, they were equally likely preceded by sounds on the same side (valid) or opposite side (invalid). Therefore, we conducted a within-subjects ANOVA including the factors Training Session and Vertical Location/Validity. As illustrated in Fig. 4, observers' ability to detect the target shape improved across sessions. They became more sensitive (d') to the shape and responded faster to it as training proceeded. This was confirmed by a significant main effect of Training Session for response times, $F(3, 48) = 6.1, p < .001$, and measures of d' , $F(3, 48) = 16.5, p < .001$. Moreover, shape detection was influenced by the location and validity of the preceding sound cue. Shapes presented at the aligned aperture location (A) were detected faster and more accurately than shapes at proximal locations (P_u/P_t). Accordingly, we observed a main effect of Vertical Location/Validity on both response times, $F(4, 64) = 4.6, p = .002$, and measures of d' , $F(4, 64) = 8.7, p < .001$.

For the trained location (P_t), targets were always preceded by valid sounds. Therefore, we estimated the missing response measures for invalid trials at P_t from the responses to invalid trials of the untrained location (P_u). Apertures were presented on different parts of the visual field (P_t vs. P_u) and it is possible that this difference had a small effect on responses. In order to deal with this problem we subtracted baseline differences, which were estimated by comparing responses to valid P_u trials with valid P_t trials at the first training session, from all estimates for the invalid P_t response measures. This procedure allowed us to separate the effects of Vertical Location and Validity. A three-way ANOVA revealed an additional three-way interaction of Training Session \times Vertical Location \times Validity on response times, $F(12, 192) = 2.0, p = .029$, but not d' measures suggesting that sounds (although irrelevant) affected shape detection, but that this crossmodal validity effect varied across vertical aperture locations and training sessions. In order to resolve this three-way interaction, we performed paired-samples t tests on the validity effect (VE) separate for each aperture location (Fig. 5). At aligned (A) locations, crossmodal facilitation (responses to validly cued shapes faster than to invalidly cued) was observed during early training (sessions 1/2), $t(16) = 2.4, p = .032$, but it disappeared as training continued. At the trained location (P_t), crossmodal facilitation was initially absent but emerged in later sessions, that is 5/6, $t(16) = 2.6, p = .020$, and 7/8, $t(16) = 2.7, p = .014$. No crossmodal facilitation was found for the untrained (P_u) control location.

Observers fixated the center of the screen fairly well. Nevertheless, we conducted a within-subjects ANOVA on eye movement measures including the factors Training Session and Vertical Location/Validity. Observers were less precisely fixating during sessions 5/6 (2.7%) than during other training sessions (.8%, 1.2%, .7%) as revealed by a main effect of Training Session on the proportion of not fixated trials, $F(3, 24) = 4.1, p = .017$. Most important, we observed no main effect or interaction with Vertical Location/Validity. We also found no effect of gaze deviation from fixation. Accordingly, the validity effects on shape detection may not be attributed to retinal shifts induced by eye movements.

Discussion

Crossmodal facilitation versus inhibition

Our findings showed that sound cues affected visual perception in two different ways: Gabor patches appearing briefly after the sound (150-ms SOA) were discriminated more accurately when the Gabors overlapped with the sound source as compared to Gabors at other visual field locations. This crossmodal facilitation is in accord with the literature showing that sound cues transiently facilitate visual discrimination (Mazza et al., 2007; McDonald et al., 2000; Spence & Driver, 1997). However, crossmodal facilitation quickly fades and even reverses to crossmodal inhibition after longer sound-light delays (Spence & Driver, 1998a, 1998b). Consistent with the literature, we observed impaired discrimination performance for Gabors appearing at the same side as the sound but with a 1,000-ms SOA. In order to test for speed-accuracy trade-offs we also examined response times. Response times did not differ significantly for validly and invalidly cued Gabor presentations and tended to be complementary to the accuracy measures (faster responses for more accurate discrimination). We recorded eye movements in order to test for the possibility that eye movements contaminated our results. We observed no difference in the pattern of eye movements for validly versus invalidly cued Gabors or for any of the tested Gabor locations suggesting that retinal shifts due to eye movements may not account for the validity effects.

At the first test session, crossmodal facilitation emerged at visual field locations that overlapped with the sound cue (A), but it was absent at visual field locations only 6° of visual angle apart (P, D). Spatially specific crossmodal facilitation has been observed before. For instance, Frassinetti et al., (2002) found that sounds facilitate luminance detection of

stimuli that were overlapping with the sound, but not of stimuli presented 16° of visual angle apart. Similarly, Meyer et al. (2005) found that sounds facilitate visual motion detection at visual field locations overlapping with the sound, but not at locations about 18° apart. These researchers only tested for horizontal auditory-visual displacements of 16° or more. Our results showed an even sharper gradient for audio-visual facilitation. The spatial resolution for cross-modal facilitation (about 6°) is comparable to the resolution of 'unimodal' auditory or visual maps. It hardly exceeds the minimum audible angle for vertical sound localization (3–5°) (Strybel & Fujimoto, 2000) or the receptive field sizes of neurons in early visual cortex (up to about 5° in the peripheral visual field) (Dumoulin & Wandell, 2008; Smith, Singh, Williams, & Greenlee, 2001). Other research has shown that crossmodal facilitation induced by sound cues affects early visual event-related potentials (McDonald et al., 2005) originating in early visual cortex (McDonald et al., 2003) and modulates early visual processing (Beer & Watanabe, 2009). These findings suggest that crossmodal facilitation likely reflects mechanisms that directly link auditory and visual maps at a relatively early processing level (Beer & Watanabe, 2009; Budinger, Heil, Hess, & Scheich, 2006; Foxe & Schroeder, 2005; Leo, Bertini, di Pellegrino, & Ladavas, 2008).

At longer SOAs (1,000 ms), crossmodal inhibition showed a relatively poor spatial specificity. At the first test session, we found no significant differences of the negative validity effect across all tested visual field locations (A, P, D). Our results, therefore, suggest that crossmodal facilitation and inhibition reflect two separate mechanisms of crossmodal interactions. Traditionally, the late inhibition effect was thought to reflect a second stage of a biphasic attention process (Klein, 2000; Maylor, 1985; Posner et al., 1985). More recently, several authors suggested that the slow inhibition effect reflects suppressed oculomotor responses towards the cued location (Collie et al., 2000; Lepsien & Pollmann, 2002; Mayer et al., 2004; Taylor & Klein, 2000). Our results show that the late inhibition effect developed slowly and affected a large proportion of the visual field. It likely involves higher-level processes. By contrast, crossmodal facilitation emerged with a short latency, faded quickly, and showed a relatively sharp spatial gradient with no noticeable effects only 6° of visual angle apart from the sound source. It likely reflects mechanisms at a relatively early processing level (Beer & Watanabe, 2009; Budinger et al., 2006; Foxe & Schroeder, 2005; Leo et al., 2008).

Multisensory perceptual learning

An important goal of our study was to examine whether crossmodal interactions are subject to multisensory perceptual learning, which would indicate experience-dependent multisensory plasticity in adult humans. Therefore, we tested for crossmodal interactions before and after multisensory perceptual learning sessions. We found that training resulted in a shift of the audio-visual tuning curve for crossmodal facilitation from the generic (aligned) to the trained (proximal) location. By contrast, training affected crossmodal inhibition only in regions remote from the trained region—effectively shrinking the range of crossmodal inhibition without changing its position. This finding shows that perceptual learning affected both mechanisms of cross-modal interactions: facilitation and inhibition. The shift of the crossmodal validity effect may not be attributed to purely unimodal perceptual learning. Sounds and Gabors were task-irrelevant throughout training sessions. Moreover, all stimuli appeared equally often at all locations. The contingency between sounds and Gabors was only established by the task-relevant target shape. Only the combination of same-side sounds and Gabors at the trained (P_T) location predicted the target shape whereas other sound-Gabor pairs were equally often paired with the target. Accordingly, changes in the crossmodal validity effect suggest a strengthening of the links between sounds and Gabors at the trained (P_T) location rather than unimodal perceptual learning.

Our findings show that multisensory perceptual learning affected both fast and slow mechanisms of crossmodal processing in adult humans. This finding seems to be inconsistent with a plethora of research that suggested that multisensory plasticity is either limited to the early life span (Bavelier & Neville, 2002; Brainard & Knudsen, 1998; King & Moore, 1991; Knudsen, 2002; Lee et al., 2001; Wallace & Stein, 2007; Wilmington et al., 1994) or to relatively late levels of multisensory processing (Fieger et al., 2006; Shinn-Cunningham, 2000; Zwiers et al., 2003). However, note that previous approaches tested for multi-sensory plasticity based on passive experience after sensory deprivation (e.g., blindness, deafness) (Fieger et al., 2006; Lee et al., 2001; Wilmington, Gray, & Jahrsdoerfer, 1994) or mis-aligned audio-visual space (e.g., by prisms or lenses) (Brainard & Knudsen, 1998; King & Moore, 1991; Knudsen, 2002; Wallace & Stein, 2007; Zwiers et al., 2003). By contrast, in our study sound-Gabor pairs—although task-irrelevant throughout training—were systematically paired with a task-relevant target stimulus. Polley et al. (2008) proposed that low-level unimodal plasticity in adults is gated by ‘reinforcement’ signals (e.g., neuromodulators) whereas plasticity during early life span only requires passive ‘exposure’ to an enriched environment. Consistent with this notion task-irrelevant low-level visual perceptual learning in adults is modulated by a simultaneously presented task-relevant target stimulus (Seitz & Watanabe, 2003) suggesting that this target stimulus gates perceptual learning by a ‘reinforcement’ signal. Alternatively, it was proposed that plasticity is prevented by inhibitory circuits and that learning only occurs when inhibition is blocked (Gutfreund, Zheng, & Knudsen, 2002). We observed crossmodal facilitation only for audio-visual associations that were paired with a relevant target shape during training (P_T) despite the fact that Gabors at other locations (A, P_U) were equally often paired with the sounds. Therefore, multisensory perceptual learning in our study likely reflects plasticity that was gated by a ‘reinforcement’ signal or by a release of inhibition rather than plasticity due to passive exposure.

The analysis of the training sessions showed that the audio-visual links responsible for crossmodal facilitation developed gradually and that learning involved two processes. Initially, crossmodal facilitation disappeared for the generically aligned (A) location. Then, crossmodal facilitation re-appeared for the trained proximal location (P_T). It is known from prism-raised barn owls that plasticity of audio-spatial maps in tectal structures involves at least two mechanisms: First, axons sprout from the central inferior colliculus—a primarily auditory structure—to the external inferior colliculus (Feldman & Knudsen, 1998). Second, connections between the external inferior colliculus and the optic tectum—a primarily visual structure—representing generic audio-visual links become inhibited by synaptic plasticity (DeBello, Feldman, & Knudsen, 2001). Interestingly, a critical period seems to apply only for the first type of multisensory plasticity as adult owls that were raised with prisms are still able to re-learn innate audiovisual tuning (Brainard & Knudsen, 1998). Axonal sprouting—a relatively slow mechanism—unlikely accounts for plasticity in the present study that involved only eight days of training. It rather seems that axonal connections between auditory and visual space representations pre-existed on a large scale but were ineffective prior to training. The shift of audio-visual spatial tuning likely reflects synaptic plasticity, that is, an inhibition of aligned connections and strengthening or dis-inhibition of trained connections.

The effect of perceptual learning on crossmodal inhibition was complex. The post versus pre-training comparison for long SOA trials in Fig. 3b seems to show a small but non-significant shift of crossmodal inhibition towards the trained location—similar to the shift of crossmodal facilitation. Similar learning effects for crossmodal facilitation and inhibition would be consistent with notions suggesting that the slow inhibition effect is the result of a single biphasic attention process (Klein, 2000; Maylor, 1985; Posner et al., 1985). Although we cannot rule out the possibility that refractoriness effects of the initial facilitation made

some contribution to crossmodal inhibition, the major effect of perceptual learning was to limit crossmodal inhibition to those visual field locations that were stimulated during training ($P_u/A/P_t$) without changing its position. This pattern of result is more consistent with an oculomotor account for the inhibition effect (Lepsien & Pollmann, 2002; Mayer et al., 2004), which predicts no spatially specific learning effects. However, a simple oculomotor account does not explain why crossmodal inhibition became weaker for those aperture locations that were not stimulated during training (D_u/D_t). This finding is puzzling and further research is needed to provide a satisfactory explanation. At present, we can only speculate: It is known that oculomotor activity—in particular saccades—can inhibit visual processing in early visual brain areas (e.g., Vallines & Greenlee, 2006). Moreover, feedback signals can modulate perceptual learning in sensory brain areas (Li, Piech, & Gilbert, 2004; Polley, Steinberg, & Merzenich, 2006). It is possible that the result of learning was to strengthen feedback connections to visual field locations that were stimulated during training while loosening feedback connections to other visual field locations. Within this framework, sounds would cause suppression of the oculomotor system. On longer SOA trials, oculomotor suppression would feed back to sensory areas. However, only feedback connections targeting sensory circuits that are activated would strengthen while all other connections would weaken. Consequently, perceptual learning effects of crossmodal inhibition would be constrained to those aperture locations that are stimulated during training.

Conclusions

In summary, sounds affected visual perception by two separate mechanisms: A fast crossmodal facilitation process that fades within less than 300 ms and that is characterized by sharp audio-visual spatial tuning and a slow inhibition process with only moderate audio-visual spatial tuning. Multisensory perceptual learning resulted in a shift of audio-visual spatial tuning for crossmodal facilitation. By contrast, training reduced the range of inhibition with no shift in audio-visual tuning. Our results suggest multisensory perceptual learning may reshape both fast *and* slow mechanisms of multisensory processing in adult humans.

Acknowledgments

This work was supported by grants from the National Institute of Health (R01 EY015980-01, R21 EY02342-01), the Human Frontier Science Program (RGP 18/2004), and the National Science Foundation (BCS-0345746, BCS-0549036). The authors declare no competing interests. We thank the reviewers for their valuable comments.

References

- Alais D, Burr D. No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research*. 2004; 19(2):185–194.10.1016/j.cogbrainres.2003.11.011 [PubMed: 15019714]
- Bauer BB. Phasor analysis of some stereophonic phenomena. *The Journal of the Acoustical Society of America*. 1961; 33(11):1536–1539.
- Bavelier D, Neville HJ. Cross-modal plasticity: Where and how? *Nature Reviews. Neuroscience*. 2002; 3(6):443–452.10.1038/nrn848
- Beer AL, Röder B. Attention to motion enhances processing of both visual and auditory stimuli: An event-related potential study. *Cognitive Brain Research*. 2004a; 18:205–225.10.1016/j.cogbrainres.2003.10.004 [PubMed: 14736579]
- Beer AL, Röder B. Unimodal and crossmodal effects of endogenous attention to visual and auditory motion. *Cognitive, Affective & Behavioral Neuroscience*. 2004b; 4(2):230–240.10.3758/CABN.4.2.230

- Beer AL, Röder B. Attending to visual or auditory motion affects perception within and across modalities: An event-related potential study. *The European Journal of Neuroscience*. 2005; 21:1116–1130.10.1111/j.1460-9568.2005.03927.x [PubMed: 15787717]
- Beer AL, Watanabe T. Specificity of auditory-guided visual perceptual learning suggests crossmodal plasticity in early visual cortex. *Experimental Brain Research*. 2009; 198:353–361.10.1007/s00221-009-1769-6
- Bower, TGR. *Development in infancy*. San Francisco, CA: Freeman; 1974.
- Brainard DH. The psychophysics toolbox. *Spatial Vision*. 1997; 10(4):433–436. [PubMed: 9176952]
- Brainard MS, Knudsen EI. Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *The Journal of Neuroscience*. 1998; 18(10):3929–3942. [PubMed: 9570820]
- Budinger E, Heil P, Hess A, Scheich H. Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience*. 2006; 143(4): 1065–1083.10.1016/j.neuroscience.2006.08.035 [PubMed: 17027173]
- Collie A, Maruff P, Yucel M, Danckert J, Currie J. Spatiotemporal distribution of facilitation and inhibition of return arising from the reflexive orienting of covert attention. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26(6):1733–1745. [PubMed: 11129370]
- DeBello WM, Feldman DE, Knudsen EI. Adaptive axonal remodeling in the midbrain auditory space map. *The Journal of Neuroscience*. 2001; 21(9):3161–3174. [PubMed: 11312301]
- Dumoulin SO, Wandell BA. Population receptive field estimates in human visual cortex. *Neuroimage*. 2008; 39(2):647–660.10.1016/j.neuroimage.2007.09.034 [PubMed: 17977024]
- Eimer M. Crossmodal links in spatial attention between vision, audition, and touch: Evidence from event-related brain potentials. *Neuropsychologia*. 2001; 39:1292–1303. [PubMed: 11566312]
- Feldman DE, Knudsen EI. Experience-dependent plasticity and the maturation of glutamatergic synapses. *Neuron*. 1998; 20(6):1067–1071. [PubMed: 9655495]
- Fieger A, Röder B, Teder-Sälejärvi W, Hillyard SA, Neville HJ. Auditory spatial tuning in late-onset blindness in humans. *Journal of Cognitive Neuroscience*. 2006; 18(2):149–157.10.1162/jocn.2006.18.2.149 [PubMed: 16494677]
- Foxe JJ, Schroeder CE. The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*. 2005; 16(5):419–423.10.1097/00001756-200504040-00001 [PubMed: 15770144]
- Frassinetti F, Bolognini N, Ladavas E. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*. 2002; 147(3):332–343.10.1007/s00221-002-1262-y
- Gibson, EJ. *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice Hall; 1969.
- Gougoux F, Zatorre RJ, Lassonde M, Voss P, Lepore F. A functional neuroimaging study of sound localization: Visual cortex activity predicts performance in early-blind individuals. *PLoS Biology*. 2005; 3(2):e27.10.1371/journal.pbio.0030027 [PubMed: 15678166]
- Grantham DW. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *The Journal of the Acoustical Society of America*. 1986; 79(6):1939–1949. [PubMed: 3722604]
- Gutfreund Y, Zheng W, Knudsen EI. Gated visual input to the central auditory system. *Science*. 2002; 297:1556–1559.10.1126/science.1073712 [PubMed: 12202831]
- King AJ, Moore DR. Plasticity of auditory maps in the brain. *Trends in Neurosciences*. 1991; 14(1): 31–37. [PubMed: 1709531]
- Klein RM. Inhibition of return. *Trends in Cognitive Sciences*. 2000; 4(4):138–146. [PubMed: 10740278]
- Knudsen EI. Instructed learning in the auditory localization pathway of the barn owl. *Nature*. 2002; 417:322–328.10.1038/417322a [PubMed: 12015612]
- Lee DS, Lee JS, Oh SH, Kim SK, Kim JW, Chung JK, et al. Cross-modal plasticity and cochlear implants. *Nature*. 2001; 409:149–150.10.1038/35051653 [PubMed: 11196628]

- Leo F, Bertini C, di Pellegrino G, Ladavas E. Multisensory integration for orienting responses in humans requires the activation of the superior colliculus. *Experimental Brain Research*. 2008; 186(1):67–77.10.1007/s00221-007-1204-9
- Lepsien J, Pollmann S. Covert reorienting and inhibition of return: an event-related fMRI study. *Journal of Cognitive Neuroscience*. 2002; 14(2):127–144.10.1162/089892902317236795 [PubMed: 11970781]
- Li W, Piech V, Gilbert CD. Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*. 2004; 7(6):651–657.10.1038/nn1255
- Loftus GR, Masson MEJ. Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*. 1994; 1(4):476–490.
- Marks, LE. The unity of the senses: Interrelations among the modalities. New York: Academic; 1978.
- Mayer AR, Seidenberg M, Dorflinger JM, Rao SM. An event-related fMRI study of exogenous orienting: Supporting evidence for the cortical basis of inhibition of return? *Journal of Cognitive Neuroscience*. 2004; 16(7):1262–1271.10.1162/0898929041920531 [PubMed: 15453978]
- Maylor, EA. Facilitatory and inhibitory components of orienting in visual space. In: Posner, MI.; Marin, OSM., editors. *Attention and performance XI*. Hillsdale, NJ: Erlbaum; 1985. p. 189-204.
- Mazza V, Turatto M, Rossi M, Umiltà C. How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia*. 2007; 45(3):514–522.10.1016/j.neuropsychologia.2006.02.010 [PubMed: 16581094]
- McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA. Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*. 2003; 15:10–19.10.1162/089892903321107783 [PubMed: 12590839]
- McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA. Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience*. 2005; 8(9):1197–1202.10.1038/nn1512
- McDonald JJ, Teder-Sälejärvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. *Nature*. 2000; 407:906–908.10.1038/35038085 [PubMed: 11057669]
- Meredith MA, Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*. 1986a; 56(3):640–662. [PubMed: 3537225]
- Meredith MA, Stein BE. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*. 1986b; 365:350–354. [PubMed: 3947999]
- Meyer GF, Wuerger SM. Cross-modal integration of auditory and visual motion signals. *NeuroReport*. 2001; 12(11):2557–2560. [PubMed: 11496148]
- Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C. Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*. 2005; 166(3–4):538–547.10.1007/s00221-005-2394-7
- Middlebrooks JC, Green DM. Sound localization by human listeners. *Annual Review of Psychology*. 1991; 42:135–159.10.1146/annurev.ps.42.020191.001031
- Polley DB, Hillock AR, Spankovich C, Popescu MV, Royal DW, Wallace MT. Development and plasticity of intra- and intersensory information processing. *Journal of the American Academy of Audiology*. 2008; 19(10):780–798.10.3766/jaaa.19.10.6 [PubMed: 19358458]
- Polley DB, Steinberg EE, Merzenich MM. Perceptual learning directs auditory cortical map reorganization through top-down influences. *The Journal of Neuroscience*. 2006; 26(18):4970–4982.10.1523/JNEUROSCI.3771-05.2006 [PubMed: 16672673]
- Posner MI, Rafal RD, Choate LS, Vaughan J. Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*. 1985; 2(3):211–228.
- Röder B, Teder-Sälejärvi W, Sterr A, Rösler F, Hillyard SA, Neville HJ. Improved auditory spatial tuning in blind humans. *Nature*. 1999; 400:162–166. [PubMed: 10408442]
- Seitz A, Watanabe T. Is subliminal learning really passive? *Nature*. 2003; 422:36.10.1038/422036a [PubMed: 12621425]
- Shinn-Cunningham B. Adapting to remapped auditory localization cues: A decision-theory model. *Perception & Psychophysics*. 2000; 62(1):33–47. [PubMed: 10703254]

- Smith AT, Singh KD, Williams AL, Greenlee MW. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*. 2001; 11:1182–1190. [PubMed: 11709489]
- Spence C, Driver J. Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*. 1997; 59(1):1–22. [PubMed: 9038403]
- Spence C, Driver J. Auditory and audiovisual inhibition of return. *Perception & Psychophysics*. 1998a; 60(1):125–139. [PubMed: 9503917]
- Spence C, Driver J. Inhibition of return following an auditory cue. The role of central reorienting events. *Experimental Brain Research*. 1998b; 118:352–360.
- Strybel TZ, Fujimoto K. Minimum audible angles in the horizontal and vertical planes: Effects of stimulus onset asynchrony and burst duration. *The Journal of the Acoustical Society of America*. 2000; 108(6):3092–3095. [PubMed: 11144604]
- Swets JA. The relative operating characteristic in psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*. 1973; 182:990–1000.10.1126/science.182.4116.990 [PubMed: 17833780]
- Tassinari G, Aglioti S, Chelazzi L, Peru A, Berlucchi G. Do peripheral non-informative cues induce early facilitation of target detection? *Vision Research*. 1994; 34(2):179–189. [PubMed: 8116277]
- Tassinari G, Berlucchi G. Covert orienting to non-informative cues: reaction time studies. *Behavioural Brain Research*. 1995; 71:101–112. [PubMed: 8747178]
- Taylor TL, Klein RM. Visual and motor effects in inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26(5):1639–1656. [PubMed: 11039490]
- Vallines I, Greenlee MW. Saccadic suppression of retinotopically localized blood oxygen level-dependent responses in human primary visual area V1. *The Journal of Neuroscience*. 2006; 26(22):5965–5969.10.1523/JNEUROSCI.0817-06.2006 [PubMed: 16738238]
- Wallace MT, Stein BE. Early experience determines how the senses will interact. *Journal of Neurophysiology*. 2007; 97(1):921–926.10.1152/jn.00497.2006 [PubMed: 16914616]
- Watanabe T, Náñez JE Sr, Koyama S, Mukai I, Liederman J, Sasaki Y. Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nature Neuroscience*. 2002; 5(10):1003–1009.10.1038/nn915
- Watanabe T, Náñez JE, Sasaki Y. Perceptual learning without perception. *Nature*. 2001; 413:844–848.10.1038/35101601 [PubMed: 11677607]
- Weeks R, Horwitz B, Aziz-Sultan A, Tian B, Wessinger CM, Cohen LG, et al. A positron emission tomographic study of auditory localization in the congenitally blind. *The Journal of Neuroscience*. 2000; 20(7):2664–2672. [PubMed: 10729347]
- Wilmington D, Gray L, Jahrsdoerfer R. Binaural processing after corrected congenital unilateral conductive hearing loss. *Hearing Research*. 1994; 74(1–2):99–114. [PubMed: 8040103]
- Zwiers MP, Van Opstal AJ, Paige GD. Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience*. 2003; 6(2):175–181.10.1038/nn999

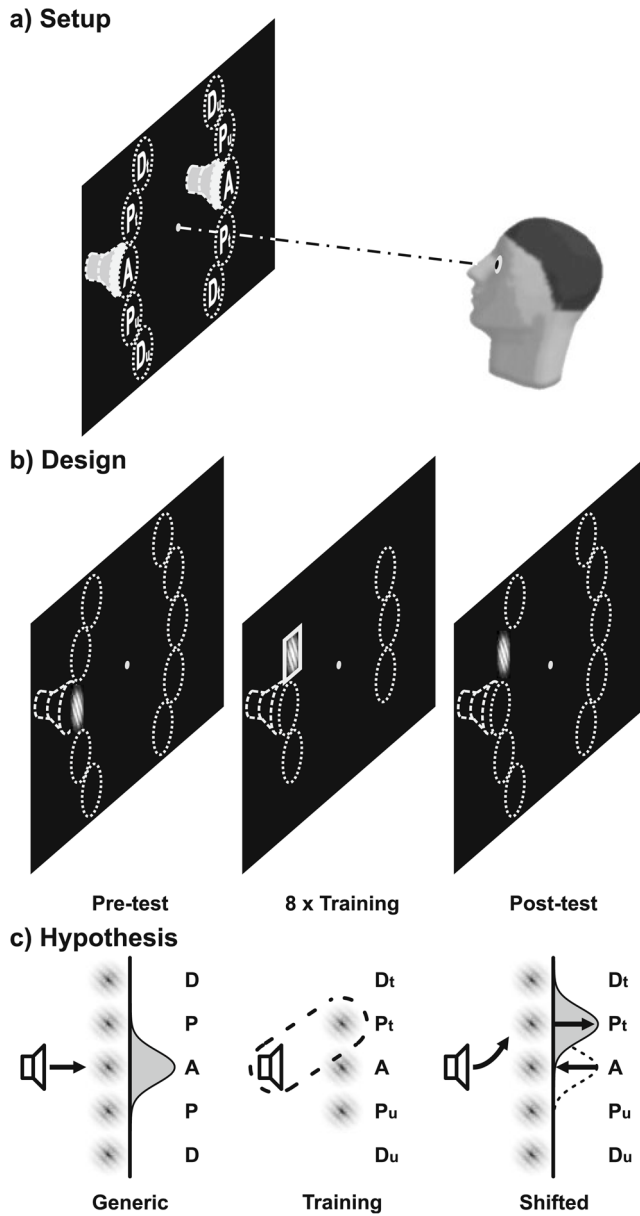


Fig. 1. Setup, design, and hypothesis. **a)** Setup. Sounds appeared to the left or right (indicated by dashed speakers). Visual field locations (dotted circles) were aligned (A), proximal (P) or distal (D) to the sound source on either side. **b)** Design. During test sessions (pre and post), crossmodal interactions were assessed for all locations ($D_t/P_t/A/P_u/D_u$) and different sound-Gabor delays (SOA: 150/300/1,000 ms). During the eight training sessions, a sound was paired with a Gabor at P_t , A, or P_u (left or right). Subsequently, a square or circle encompassed the Gabor. Target shapes established a contingency between same-side sounds and Gabors at trained locations (P_t) only. The distal locations next to P_t (D_t) or next to P_u (D_u) were not stimulated. **c)** Crossmodal facilitation at the generic locations (A) prior to training was expected to shift to the trained locations (P_t) after training (only one side depicted)

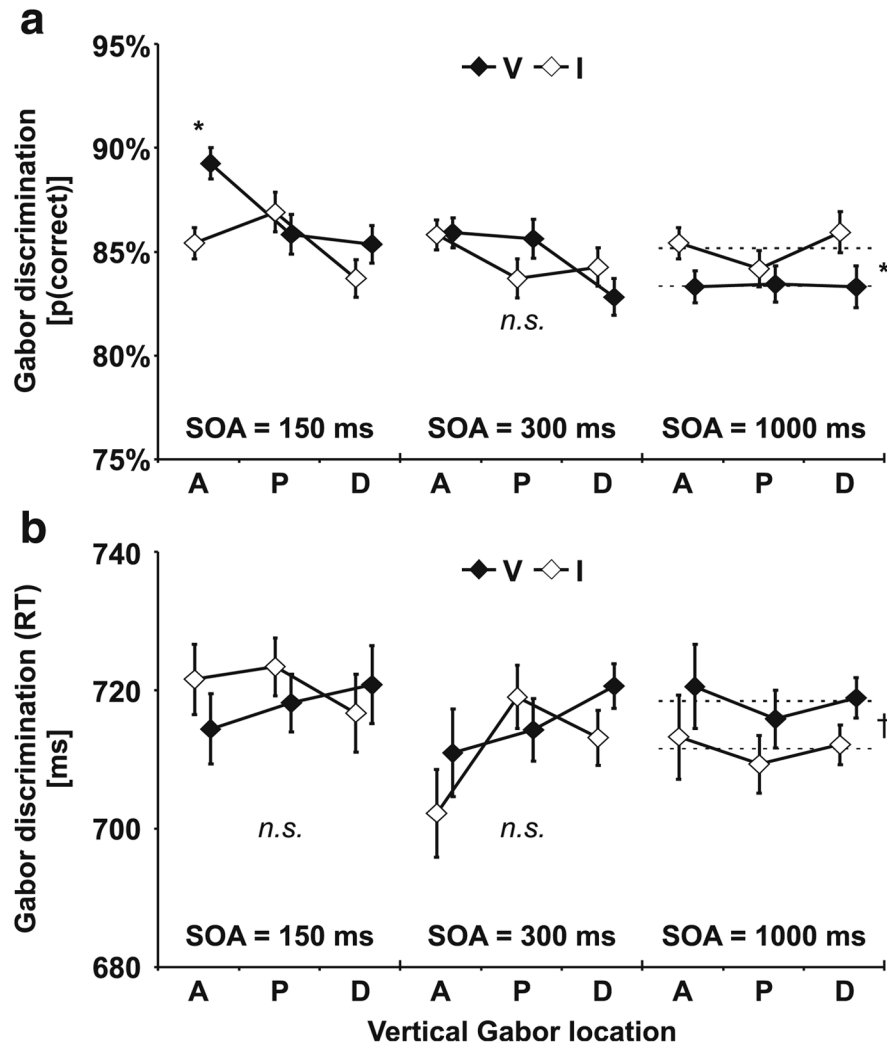


Fig. 2. Test results prior to training. **a)** Percentage correct Gabor discrimination. For the 150-ms stimulus onset asynchrony (SOA), Gabors were discriminated more correctly when preceded by valid (V) than by invalid (I) sound cues. This crossmodal facilitation was observed only at aligned (A) but not at proximal (P) or distal (D) locations. No significant (*n.s.*) differences emerged for the 300-ms SOA. For the 1,000-ms SOA, crossmodal inhibition was observed across all visual field locations (A, P, D). Dashed lines depict the mean across all aperture locations. See ANOVA results in text. **b)** No significant differences were observed on response times. However, response time differences tended to complement validity effects on percentage correct. Upper and lower visual field representations of proximal (P) and distal (D) apertures were averaged for illustrative purposes. They were separately analyzed in statistical tests (see text). *Error bars* reflect within-subjects SEM (Loftus & Masson, 1994) for the factor Validity. *n* = 12. * *p* ≤ .05. † *p* ≤ .10

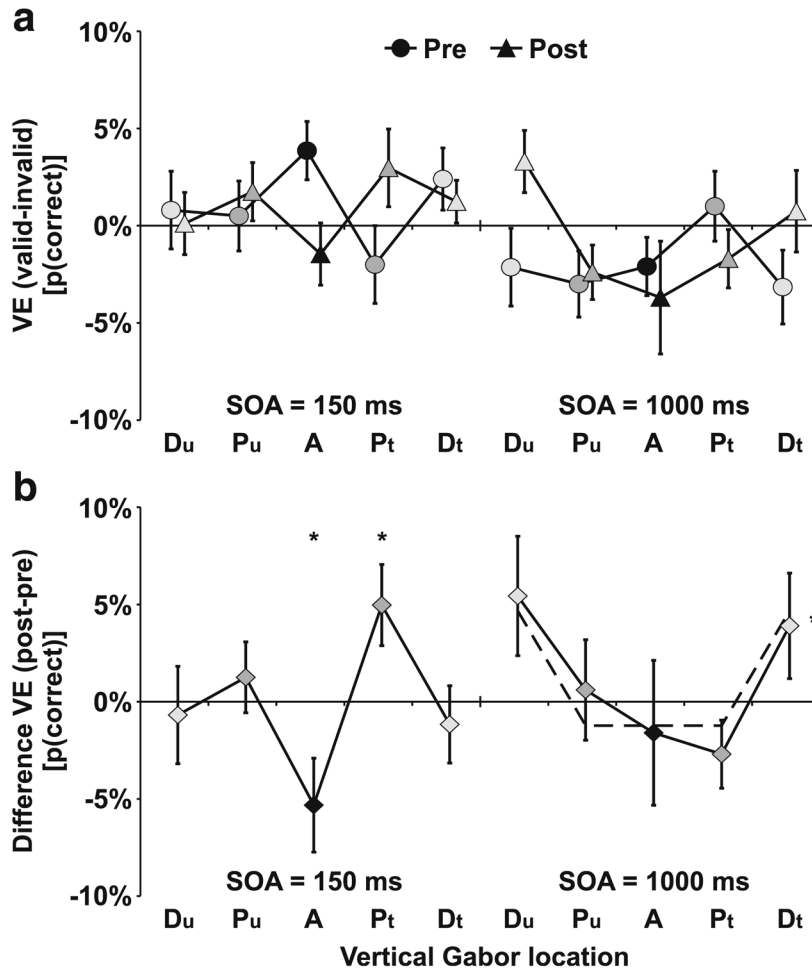


Fig. 3. Validity effects (VE). **a**) Pre- and post-training tests. Prior to training, a VE (performance at same/valid side as sound minus performance at opposite/invalid side) showing crossmodal facilitation emerged only at the aligned (A) location for the 150-ms stimulus onset asynchrony (SOA). Crossmodal inhibition was observed at the 1,000-ms SOA with no significant differences across vertical locations. *Error bars* reflect within-subjects SEM (Loftus & Masson, 1994) for the VE. **b**) Change in VE (post-training minus pre-training). Crossmodal facilitation decreased at the aligned (A) and increased at the trained location (P_t). Crossmodal inhibition decreased at distal (D_t/D_u) locations. No significant change was observed at aligned or proximal locations (P_t/A/P_u). Dashed lines depict the mean for both distal (D_t/D_u) locations and the mean for aligned or proximal locations (P_t/A/P_u). *Error bars* reflect within-subjects SEM (Loftus & Masson, 1994) for the difference in the VE. *n* = 12. * *p* ≤ .05

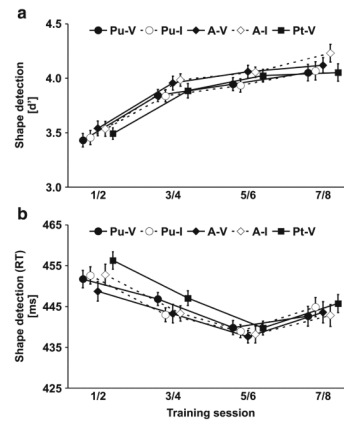


Fig. 4. Shape detection during training. **a)** Sensitivity measures (d'). **b)** Mean response times. Shape detection improved at all visual field locations (P_t , A, P_u) and both when sounds were presented at the same (valid) side (V) or the opposite (invalid) side (I). In addition, a location-specific change of the validity effect was observed on response times (see Fig. 5). Two consecutive training sessions were pooled. Error bars reflect within-subjects s.e.m. (Loftus & Masson, 1994) for the factor Training Session. $n = 17$

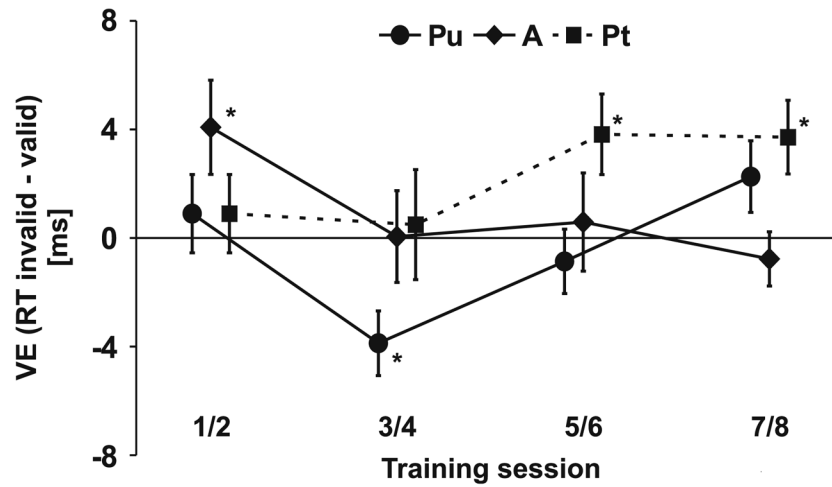


Fig. 5. Change of validity effect during training. During early training, crossmodal facilitation (VE = response times at opposite/invalid side as sound minus response times at same/valid side) only emerged at aligned (A) locations. At later sessions, crossmodal facilitation was restricted to trained locations (Pt). Error bars reflect within-subjects SEM (Loftus & Masson, 1994) for the VE. $n = 17$. * $p \leq .05$