



Published in final edited form as:

Exp Brain Res. 2009 September ; 198(2-3): 353–361. doi:10.1007/s00221-009-1769-6.

Specificity of auditory-guided visual perceptual learning suggests crossmodal plasticity in early visual cortex

Anton L. Beer and Takeo Watanabe

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

Abstract

Sounds modulate visual perception. Blind humans show altered brain activity in early visual cortex. However, it is still unclear whether crossmodal activity in visual cortex results from unspecific top-down feedback, a lack of visual input, or genuinely reflects crossmodal interactions at early sensory levels. We examined how sounds affect visual perceptual learning in sighted adults. Visual motion discrimination was tested prior to and following eight sessions in which observers were exposed to irrelevant moving dots while detecting sounds. After training, visual discrimination improved more strongly for motion directions that were paired with a relevant sound during training than for other directions. Crossmodal learning was limited to visual field locations that overlapped with the sound source and was little affected by attention. The specificity and automatic nature of these learning effects suggest that sounds automatically guide visual plasticity at a relatively early level of processing.

Keywords

Crossmodal plasticity; Multisensory integration; Perceptual learning; Visual perception; Psychophysics

Introduction

Sounds modulate visual perception (Mazza et al. 2007; McDonald et al. 2000; Spence and Driver 1997) and affect early event-related potentials (Beer and Röder 2005; Eimer 2001; McDonald et al. 2005; Teder-Sälejärvi et al. 1999) originating in early visual cortex (McDonald et al. 2003). The crossmodal interactions observed in ‘modality-specific’ areas have been ascribed to feedback projections from multimodal stages (Macaluso et al. 2000; McDonald et al. 2003, 2005). Namely, sensory signals first converge at ‘high-level’ multimodal stages before they affect signal processing of other modalities.

Congenitally blind people, when compared with sighted, showed enhanced activity in occipital brain areas including the primary visual cortex while analyzing speech (Röder et al. 2002), perceiving auditory motion (Poirier et al. 2006), or localizing sounds (Weeks et al. 2000) as revealed by functional magnetic resonance imaging. These results imply that sounds modulate neural circuits that are usually responsible for relatively early levels of visual processing. However, the activity in early sensory cortex of blind humans may equally well reflect plasticity in ‘high-level’ multimodal stages (Büchel et al. 1998), which recruit early sensory cortex by

feedback loops. Another possibility is that auditory experience in blind humans only alters auditory circuits within occipital cortex rather than affecting visual circuits. Although this option is rarely considered it cannot be ruled out based on the previous brain imaging studies. Moreover, visual experience during early life substantially determines the functional development of the visual cortex. Consequently, alterations in early visual cortex of visually deprived humans may result from a lack of visual input (Katz and Shatz 1996) or less efficient inhibitory circuits (Rozas et al. 2001) rather than reflecting low-level crossmodal interactions. Therefore, it still remains controversial whether crossmodal recruitment of early visual cortex results from unspecific top-down feedback, a lack of visual input, or genuinely reflects crossmodal plasticity.

In order to address this controversy, we examined the location and feature specificity of auditory-guided changes in visual perception by means of task-irrelevant perceptual learning. Perceptual learning (PL) refers to a form of implicit learning by which repeated sensory experience may lead to long-lasting alterations in perception (Gibson 1963). Visual PL effects are specific to the trained retinal location and simple features such as orientation or motion direction (Fiorentini and Berardi 1980; Schoups et al. 1995). This location and feature specificity suggests that the neural substrate for PL involves early visual cortex. Recent brain imaging studies in humans (Schwartz et al. 2002) and single cell recordings in monkeys (Li et al. 2004; Schoups et al. 2001) conformed that unimodal PL alters activity in primary visual cortex. As uncontrolled plasticity at low-level sensory stages might destabilize our perceptual system some researchers proposed that perceptual learning occurs only with relevant (or attended) stimulus features (Ahissar and Hochstein 1993). Moreover, the reverse hierarchy theory proposes that perceptual learning for discriminating relevant stimulus features first involves higher-level processing stages before it affects low-level sensory stages (Ahissar and Hochstein 2004). In contrast to this notion, PL can occur even when stimuli are irrelevant for the observer. This so-called 'task-irrelevant' PL does not require attention and may occur even with subliminal stimuli (Watanabe et al. 2001). Interestingly, task-irrelevant PL improves discrimination of low-level features (e.g., local visual motion) without improving discrimination of high-level features (e.g., global visual motion) suggesting that task-irrelevant PL only affects early sensory cortex (Watanabe et al. 2002).

We reasoned that if crossmodal plasticity occurs at multimodal brain areas such as parietal or temporal lobe whose neurons have large receptive fields and represent space in distributed code (Andersen et al. 1997), then crossmodal learning should affect a large proportion of the visual field. In contrast, if crossmodal plasticity affects early visual areas with small receptive fields, then crossmodal learning effects should emerge only at visual field locations that overlap with the sound source during training and should be specific to the visual feature that was paired with the sound. Voluntary (endogenous) attention to a sound source is known to spread across modalities (Spence and Driver 1996) and likely involves higher-level brain areas (Eimer 2001). We were interested in low-level plasticity. Therefore, we tested whether sounds modulate visual perception when the visual stimuli were irrelevant during the exposure. Task-irrelevant PL occurs relatively automatically. Accordingly, we expected that sounds induce changes in visual perception even when the sound location has to be ignored.

In order to examine the location and feature specificity of crossmodal learning, we tested direction discrimination performance for moving dots prior to and following eight auditory (crossmodal) training sessions at ten peripheral visual field locations: left or right, aligned (A) with the virtual sound source during training, proximal (P), or distal (D) to this aligned location (Fig. 1a, b). During auditory training (Fig. 1c), sounds were presented either at the left or right together with moving dots either at the same or opposite side than the sound. Observers were asked to detect a deviant sound among a stream of standard sounds at one side (L+) and to ignore this deviant sound when presented at the opposite side (L-). They were asked to

disregard the moving dots, which moved equally likely toward one of four directions. However, one motion direction (Dir-L+) was presented frequently together with the deviant sound at the attended location (L+), whereas a perpendicular direction (Dir-L-) was paired with the deviant sound at the ignored side (L-). We were interested in how the deviant sounds during training affect visual motion discrimination performance after training.

Materials and methods

Participants

Participants were 16 paid (\$90 total) students with normal hearing and normal or corrected-to-normal vision who volunteered after giving written informed consent. The study was approved by the Institutional Review Board of Boston University. Three participants quit before finishing all sessions and the data of one participant was excluded because of ceiling performance (>95% correct) in the first test session. The mean age of the remaining 12 participants (one male, one left-handed) was 20.9 years (18–31).

Apparatus

Participants sat in front of a CRT-monitor (40 × 30 cm, 1,280 × 1,024 pixel, 75 Hz) in a dark room. Two small speakers were mounted to the left and right of the monitor vertically aligned with the screen center (Fig. 1a). The monitor and speaker chassis as well as the adjacent space were masked black in order to prevent observers from seeing the speakers during the experiment. Participants' head (viewing distance 60 cm) was stabilized by a chin rest and a nose clamp. Eye movements were recorded (calibration window 90% of screen, 30 Hz digitization rate) from the right eye of eight participants with a View Point Quick-Clamp Camera System (Arrington Research, Scottsdale, AZ). 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 (no filtering was applied). Stimuli were presented with Psychophysics Toolbox (Brainard 1997) version 2.55 and MatLab 5.2.1 (MathWorks, Natick, MA) on a G4 Mac with OS 9.

Stimuli

Visual stimuli were white dots ($0.1^\circ \times 0.1^\circ$, 8 dots/deg², 3.38 cd/m²) randomly distributed on a black background (0.01 cd/m²) moving (8°/s) within a circular aperture (6° in diameter). Dots moved either in a random direction or coherently (30%) in one of four orthogonal directions (45°, 135°, 225°, or 315°, see Fig. 1b). The moving dots were shown for a total of 200 ms—similar to the previous studies (e.g., Watanabe et al. 2002). The direction of individual dots was randomly reassigned every 83 ms.

Auditory stimuli consisted of high or low pitch sounds. Sound pressure level at ear position was about 60 dB (A). Several aspects that likely have an impact on crossmodal learning effects were considered: first, the sounds should be distinct with regard to the features pitch and location that were expected to affect crossmodal learning. Second, the sounds should be difficult to discriminate in order to capture voluntary (endogenous) spatial attention. Finally, the sounds should be as similar as possible with regard to any other physical aspect (e.g., intensity) in order to avoid that physical saliency affects crossmodal learning. In order to meet these criteria, distinct sound frequencies were chosen for high (1,568 Hz) and low (1,046 Hz) pitch sounds. Pure tones (sine wave sounds) were embedded in white noise as vertical localization primarily relies on spectral cues (Middlebrooks and Green 1991). The sounds were only presented for a brief duration (50 ms, 2 ms rise and fall time) and the noise level was set to be slightly above the discrimination threshold as determined for each individual by a simple staircase procedure prior to training. The mean signal-to-noise ratio (average power of pure tone to white noise) was 7.7%. Finally, the average power of low and high pitch sounds was

matched. As close spatial alignment is relevant for multisensory integration (Meredith and Stein 1986), particularly in our experiment, the interaural level difference was adjusted according to the law of sines (Grantham 1986) in order to accomplish a close overlap of sounds and visual stimuli.

Design

An initial visual test session (pre) was followed by eight auditory (crossmodal) training sessions and another visual test session (post). Each session lasted about 1 h and was conducted on a separate day.

During visual test sessions dots were presented either at the left or right side at one of five vertical locations (A, aligned with virtual sound source during training; P, proximal bottom or top; D, distal bottom or top). The centers of two neighboring apertures were 6° apart from each other. Note that this distance is somewhat larger than the minimum audible angle for vertical sound localization (3°–5°) (Strybel and Fujimoto 2000). The centers of all apertures were 16° distant to fixation.

During auditory (crossmodal) training, a high or low pitch sound was presented either on the left or right. Previous research has shown that PL is modulated by the simultaneous presentation of a target stimulus (Seitz and Watanabe 2003). Participants had to detect deviant sounds (e.g., high pitch sound) presented at one side only (L+, left or right counterbalanced across participants) and to ignore distractor sounds (e.g., low pitch sound) or sounds at the irrelevant side (L–). The deviant sound alternated every session and was presented 20% of the trials. Concomitantly with the sounds, a dot pattern was presented aligned with the relevant (L+) or irrelevant sound source (L–). Combinations of sound and dot location were balanced. One direction (Dir-L+) of the coherently moving dots on either side was presented more frequently (75%) together with the deviant sound at the relevant side (L+). Another direction (perpendicular to Dir-L+) was paired (75%) with the deviant sound at the irrelevant side (L–). The two remaining distractor directions were equally tied to left or right distractor sounds.

Procedure

During visual test trials, moving dots were presented for 200 ms at one of the ten locations. Then, dot patterns disappeared and after a delay of 100 ms participants were asked to indicate in which direction the coherently moving dots were heading by pressing one of four keys. The inter-trial interval varied randomly between 500 and 900 ms. Participants were asked to fixate a central bull's eye throughout the experiment. A test session consisted of four blocks of 400 trials. No feedback was provided during or after these blocks. A brief practice block (40 trials) with increased coherence level (80%) was run in order to familiarize participants with the task.

Trials for the auditory (crossmodal) training started with the presentation of moving dots at one of the two locations closest to the speaker (A, aligned). After 100 ms a sound was presented for 50 ms at the same or the opposite side as the visual stimulus. The moving dots disappeared 50 ms after sound offset. This timing was chosen in order to assure that the moving dots were presented for the whole sound duration. The inter-trial interval was 600 ms. Participants' task was to press the space bar whenever they heard a target sound coming from the relevant side (L+) and to ignore all other sounds. They were asked to ignore the moving dots and to fixate a central bull's eye throughout the experiment. A training session consisted of four blocks of 1,024 trials each. Performance feedback was provided after each block.

Results

Tests

At the first test session, participants correctly detected the motion direction on 69.9% of the trials. No significant differences on detection performance were found between the four motion directions, the vertical aperture locations (A, P, D) or between the left or right aperture side. After training, mean detection performance was 73.6%. A four-way within-subjects ANOVA revealed a significant main effect of test time (post vs. pre), $F(1, 11) = 8.4$, $P = 0.015$, as well as significant interactions of test time \times direction \times vertical location, $F(6, 66) = 3.1$, $P = 0.009$, and test time \times direction \times vertical location \times side, $F(6, 66) = 3.1$, $P = 0.009$.

In order to reveal auditory-guided changes in visual perception, we compared changes in motion discrimination (correct responses post – pre-training) for directions that were paired with the deviant sound (Dir-L+, Dir-L-) at locations that were spatially aligned with the sounds during training. As illustrated in Fig. 2a and conformed by a within-subjects ANOVA showing an interaction of side by direction, $F(1, 11) = 22.7$, $P = 0.001$, our training procedure resulted in better discrimination at the attended side (L+) than at the opposite side (L-) for the direction tied with the deviant sound at the attended side (Dir-L+), $t(11) = 3.5$, $P = 0.005$. In contrast, visual discrimination improved more at the ignored side (L-) than at the attended side (L+) for the motion direction that was paired with deviant sounds at the ignored side (Dir-L-), $t(11) = 2.3$, $P = 0.038$. A similar pattern of results emerged when sensitivity measures (d') (Swets 1973) rather than percentage correct were analyzed. Although responses at the post-test were faster than at the pre-test (Fig. 3), $F(1, 11) = 8.1$, $P = 0.016$, no location or direction differences were observed on response times arguing against speed-accuracy trade-offs. Trials with eye movements (see “Apparatus” for recording procedure) during stimulus presentation were rare (overall 2.3% of the trials). The proportion of trials with eye movements did not differ significantly for each location or motion direction condition.

We further tested for the spatial gradient of the crossmodal learning effect. As illustrated in Fig. 2b and conformed by a main effect of vertical location, $F(2, 22) = 8.1$, $P = 0.002$, auditory-guided changes in visual discrimination (absolute difference in learning between L+ and L- separate for each trained direction and vertical level) was only observed at locations overlapping with the sound source (A) with no effect at locations that were at the same 16° eccentricity but 6° above or below (P, D). As mentioned above, we were further interested in whether voluntary attention during training had an effect on crossmodal learning. Therefore, we compared the crossmodal learning effect for directions paired with sounds at the attended location (Dir-L+) and directions paired with sounds at the ignored side (Dir-L-). A within-subject ANOVA indicated no difference ($P = 0.721$) between these two directions. Eye movements (proportion of not fixated trials) did not differ ($P > 0.1$) across aperture locations or across motion directions.

It may be argued that the sharp spatial gradient of our crossmodal learning effects result from intrinsic properties of the horizontal meridian of the visual cortex. Therefore, we compared unimodal learning effects on distractor directions (Fig. 2c, 3). Observers identified these distractor directions (Dir-D1/2) more correctly, $F(1, 11) = 7.1$, $P = 0.022$, and faster, $F(1, 11) = 15.8$, $P = 0.002$, at the post-test than at the pre-test. This learning effect was observed at all vertical aperture locations with no difference between left or right apertures. No differences across conditions were observed on eye movements.

Training

In order to find out what caused crossmodal learning, we analyzed auditory performance during training sessions. Response times to target sounds (Fig. 4a) decreased across sessions, $F(7, 77)$

= 24.1, $P < 0.001$, reflecting unimodal auditory learning. Moreover, responses were faster when the sound was presented at the same location as the visual stimulus (dots at L+) when compared with the opposite side (L-), $F(1, 11) = 32.8$, $P < 0.001$. The proportion of correctly detected sounds (hits) increased across sessions, $F(7, 77) = 6.7$, $P = 0.025$, and was higher when the sounds were presented at the same location as the dots than at the opposite side (e.g., 66.7 vs. 57.5% at first training session), $F(1, 11) = 2.2$, $P = 0.041$. Higher hit rates were also associated with higher error rates (false alarms). False alarms for sounds at the irrelevant side (L-) but of the relevant pitch (distractors) were higher for sounds with spatially congruent (11.2%) as compared to incongruent dots (4.8%), $F(1, 11) = 6.6$, $P = 0.026$. False alarm rates for sounds of the irrelevant pitch were higher for sounds at the relevant (7.3% at L+) than at the irrelevant side (1.7% at L-), $F(1, 11) = 5.3$, $P = 0.041$.

Faster responses and higher hit rates to spatially congruent audio-visual stimuli may either reflect multisensory enhancement at a sensory level or response priming. In order to distinguish between these two alternatives sensitivity measures (d') and measures of response bias (β) (Swets 1973) were calculated. Sounds had to be detected based on two criteria: location and pitch. Therefore, d' and β were calculated separately for each of these stimulus dimensions. Spatial discrimination measures were calculated based on the false alarm rates for sounds of the relevant pitch but at the irrelevant side. Pitch discrimination measures were calculated based on the false alarm rates for sounds at the relevant side but of the irrelevant pitch. Note that this approach tends to overestimate d' as hit rates were based on stimuli complying with both relevant stimulus dimensions whereas false alarm rates were based on the stimuli that violate only one stimulus dimension. However, by assuming that pitch and location information aggregates independently, this approach allows us to estimate the relative discriminability for each stimulus dimension. Sensitivity (Fig. 4b) for spatial discrimination increased across sessions, $F(7, 77) = 2.6$, $P = 0.019$. Interestingly, d' values were lower when the target sound was presented together with the visual stimulus at the same location (dots at L+) rather than the opposite side (L-), $F(1, 11) = 8.5$, $P = 0.014$, suggesting crossmodal interference rather than enhancement. No significant differences in sensitivity were found for pitch discrimination.

Measures of response bias (β) are depicted in Fig. 4c. The tendency to respond to irrelevant sounds increased across sessions for spatial discrimination, $F(7, 77) = 3.6$, $P = 0.002$, and for pitch discrimination, $F(7, 77) = 2.8$, $P = 0.011$. Moreover, this response bias was stronger when the dots were presented at the same side as the sound for both spatial, $F(1, 11) = 6.4$, $P = 0.028$, and pitch discrimination, $F(1, 11) = 6.9$, $p = 0.024$.

Eye movements (proportion not fixated trials) tended to increase across training sessions (3.5% at first, 6.1% at last session) as indicated by a marginally significant main effect of session, $F(7, 49) = 2.1$, $P = 0.061$. However, eye movements did not differ ($P > 0.6$) for left and right sounds, left and right visual apertures, or for congruent (dots at L+) or incongruent (dots at L-) trials.

Discussion

Visual discrimination performance for motion directions that were tied with a deviant sound (Dir-L+/L-) during training improved more at visual field locations that overlapped with the associated sound than at the opposite side. As both motion directions were equally often presented on both sides and were equally likely associated with the sounds this finding may not be attributed to pure exposure to moving dots or to an unspecific alertness or learning signal (Seitz et al. 2006). Instead, it reflects crossmodal learning that is specific to the motion direction that was paired with the sound. This finding demonstrates that sounds not only affect instantaneous visual perception (Beer and Röder 2005; Driver and Noesselt 2008; Eimer

2001; Mazza et al. 2007; McDonald et al. 2000; McDonald et al. 2005; Spence and Driver 1997), but also guide visual perceptual learning.

The auditory-guided learning effects on visual perception were similar regardless of whether the sound source was attended (Dir-L+) or ignored (Dir-L-) during training. This finding is in-line with research showing that low-level visual perceptual learning does not require attention (Watanabe et al. 2001, 2002). As attending to the sound source yielded no measurable additional benefit on visual learning, crossmodally guided perceptual learning seemed to occur rather automatically without the need of voluntary (endogenous) attention (e.g., Eimer 2001; Spence and Driver 1996).

It might be argued that participants divided their attention equally to both the relevant and the irrelevant sound location rather than attending to one side only. Although we cannot exclude the possibility that observers occasionally attended to sounds from the to-be-ignored side it likely had only a marginal effect on visual learning. We found that false alarm rates for pitch discrimination were substantially higher at the relevant side than at the irrelevant side suggesting that participants allocated much more attention to the relevant sound location. Moreover, if voluntary attention to the sound location would have been sufficient to guide visual perceptual learning our learning effects should be equal for all motion directions—rather than direction specific.

The crossmodal learning effects in our study were only observed at visual field locations that were aligned (A) with the sound source during training but were absent at closeby locations only 6° of visual angle apart (P or D). The relatively sharp spatial gradient of our crossmodal learning effects resemble the location specificity (Crist et al. 1997; Karni and Sagi 1991; Schoups et al. 1995; Sowden et al. 2002) of low-level visual perceptual learning (Schoups et al. 2001; Schwartz et al. 2002; Watanabe et al. 2001, 2002). Receptive fields in early visual areas only cover a few degrees of visual angle (Dumoulin and Wandell 2008; Smith et al. 2001), whereas receptive fields of higher-level areas such as MT (Felleman and Kaas 1984) or parietal lobe (Andersen et al. 1997) represent large proportions of the visual scene (about 25°–40°). Accordingly, low-level perceptual learning is expected to be specific to the location of the training stimuli and should not transfer to other visual field regions.

Our crossmodal learning effects not only were constrained by the location of the sounds but were also specific to the motion direction that was paired with the sound. Again, this is consistent with the previous findings showing that low-level visual perceptual learning is specific to the trained feature such as orientation and motion direction (Crist et al. 1997; Watanabe et al. 2001, 2002). Direction-selective cells are found in several low-level (e.g., V1) and mid- or high-level visual areas (e.g., MT). Neurons in V1 are tuned to only one motion direction, whereas neurons in MT encode global motion and receive afferent signals for different motion directions (Heeger et al. 1999; Xiao et al. 1997). Accordingly, low-level perceptual learning should be feature-specific, whereas high-level learning should show a substantial degree of transfer to other motion directions.

During training, moving dots were only presented at aligned (A) locations along the horizontal meridian. It may be objected that learning was constrained by intrinsic properties of the horizontal meridian rather than by the alignment of the sounds. However, unimodal perceptual learning was observed at all tested visual field locations suggesting that learning was possible across all aperture locations. Furthermore, the lack of transfer to vertical aperture locations not overlapping with the sounds (P, D) might be attributed to the lack of visual stimulation at these aperture locations. This notion would not contradict our conclusion that learning occurred at early sensory levels as due to the larger receptive fields of higher-level areas learning effects should transfer to visual field locations that were not directly stimulated. Moreover, pure

exposure to visual stimuli may not explain our results since we observed learning effects that were specific to the motion direction despite all directions were equally often presented during training. Instead our data suggests that the sharp spatial gradient for crossmodal learning effects arose because learning was constrained by the alignment of the sounds with the moving dots during training.

Consistent with the previous research (Mazza et al. 2007; Spence and Driver 1997) our analysis of sensitivity measures (d') revealed that the dot patterns presented during the training were not facilitating auditory perception. This finding rules out the possibility that sounds presented at the same location as the moving dots were perceived more vividly and induced learning by raising alertness or by means of an unspecific learning signal (Seitz et al. 2006). As we observed faster responses and a criterion shift (β) when sounds and dots were aligned it could be argued that motor preparation and/or execution guided perceptual learning. According to this notion stronger learning would be expected at the attended side (L+) than at the ignored side (L-). In contrast, crossmodal effects were about equally strong for directions tied to either the relevant or irrelevant sound location, respectively, suggesting that this response bias had no substantial effect.

Interestingly, sensitivity (d') for sounds was reduced, rather than enhanced, by spatially aligned sounds. Although such crossmodal interference has been observed before (Mazza et al. 2007) it is not yet fully understood. It may be the paradoxical result of strong multisensory integration. For instance, sounds automatically enhance the saliency of spatially aligned visual stimuli (Mazza et al. 2007; McDonald et al. 2000; Spence and Driver 1997). Furthermore, irrelevant moving dots interfere with the perception of relevant visual targets and trigger a feedback mechanism that suppresses sensory signals (Tsushima et al. 2006). More salient visual distractors would elicit stronger inhibition. Nevertheless, inhibition cannot explain the crossmodal learning effects in our study as weaker visual signals due to inhibition should induce weaker learning and there should be either no difference between left and right locations or weaker learning at locations spatially aligned with the sound source—the opposite was observed.

The most likely explanation of our crossmodal learning effects is that sounds during training either enhanced the saliency of spatially aligned irrelevant moving dots (Mazza et al. 2007; McDonald et al. 2000; Spence and Driver 1997) or instructed learning specifically for these stimuli. The deviant and distractor sounds were about equally salient based on physical attributes. They were matched in average power and the target pitch was alternated each training session. However, deviants were less likely than distractor sounds and electrophysiology has shown that rare auditory events automatically (Rinne et al. 2001) trigger a mismatch-negativity response in auditory cortex (Jääskeläinen et al. 2004). This auditory change-detection signal may serve as the signal needed for instructing learning in vision. It would be interesting to see in future research how change-detection (e.g., the degree of deviance) affects unimodal and crossmodal perceptual learning.

Finally, it might be argued that change-detection made the sounds psychologically salient and this saliency attracted attention to the moving dots via top-down control. This notion involves that stimulus-driven (exogenous) attention affects stimulus processing in higher-level multimodal brain regions (Macaluso et al. 2000; McDonald et al. 2003, 2005). From the literature on perceptual learning (e.g., Ahissar and Hochstein 2004) we know that if a task requires the involvement of higher-level brain areas there should be a relatively strong transfer across stimulus locations and features. In contrast, we found little crossmodal learning for aperture locations or motion directions that were not overlapping and paired with the deviant sound.

In summary, our findings suggest that sounds guided visual perceptual learning rather automatically. Crossmodal learning was specific to the location and visual feature that was paired with the sound. Our results are consistent with several studies showing that sounds modulate brain activity in visual cortex of blind humans more strongly than in sighted (Gougoux et al. 2005; Poirier et al. 2006; Röder et al. 2002; Weeks et al. 2000). Our results further demonstrate that auditory processing not only co-occurs with visual processing but instead that sounds induce long-lasting changes in visual perception. The automaticity and specificity of our crossmodal learning effects suggest that sounds directly alter visual perception at relatively early sensory levels. Crossmodal plasticity in early sensory cortex does not require top-down feedback (Büchel et al. 1998; Macaluso et al. 2000; McDonald et al. 2003, 2005) or depend on a lack of visual input (Katz and Shatz 1996; Rozas et al. 2001).

Acknowledgments

This work was supported by the National Science Foundation (BCS-0549036), the National Institute of Health (R01 EY015980-01, R21 EY02342-01) and the Human Frontier Foundation (RGP18/2004).

References

- Ahissar M, Hochstein S. Attentional control of early perceptual learning. *Proc Natl Acad Sci USA* 1993;90:5718–5722. [PubMed: 8516322]
- Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 2004;8:457–464. [PubMed: 15450510]
- Andersen RA, Snyder LH, Bradley DC, Xong J. Multimodal representation of space in the posterior cortex and its use in planning movements. *Annu Rev Neurosci* 1997;20:303–330. [PubMed: 9056716]
- Beer AL, Röder B. Attending to visual or auditory motion affects perception within and across modalities: an event-related potential study. *Eur J Neurosci* 2005;21:1116–1130. [PubMed: 15787717]
- Brainard DH. The psychophysics toolbox. *Spat Vis* 1997;10:433–436. [PubMed: 9176952]
- Büchel C, Price C, Frackowiak RSJ, Friston K. Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain* 1998;121:409–419. [PubMed: 9549517]
- Crist RE, Kapadia MK, Westheimer G, Gilbert CD. Perceptual learning of spatial localization: specificity for orientation, position, and context. *J Neurophysiol* 1997;78:2889–2894. [PubMed: 9405509]
- Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 2008;57:11–23. [PubMed: 18184561]
- Dumoulin SO, Wandell BA. Population receptive field estimates in human visual cortex. *Neuroimage* 2008;39:647–660. [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]
- Felleman DJ, Kaas JH. Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *J Neurophysiol* 1984;52:488–513. [PubMed: 6481441]
- Fiorntini A, Berardi N. Perceptual learning specific for orientation and spatial frequency. *Nature* 1980;287:43–44. [PubMed: 7412873]
- Gibson EJ. Perceptual learning. *Annu Rev Psychol* 1963;14:29–56. [PubMed: 13947716]
- 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 Biol* 2005;3:e27. [PubMed: 15678166]
- Grantham DW. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J Acoust Soc Am* 1986;79:1939–1949. [PubMed: 3722604]
- Heeger DJ, Boynton GM, Demb JB, Seidemann E, Newsome WT. Motion opponency in visual cortex. *J Neurosci* 1999;19:7162–7174. [PubMed: 10436069]
- Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin F-H, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW. Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci USA* 2004;101:6809–6814. [PubMed: 15096618]

- Karni A, Sagi D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc Natl Acad Sci USA* 1991;88:4966–4970. [PubMed: 2052578]
- Katz LC, Shatz CJ. Synaptic activity and the construction of cortical circuits. *Science* 1996;274:1133–1138. [PubMed: 8895456]
- Li W, Piech V, Gilbert CD. Perceptual learning and top-down influences in primary visual cortex. *Nat Neurosci* 2004;7:651–657. [PubMed: 15156149]
- Loftus GR, Masson MEJ. Using confidence intervals in within-subject designs. *Psychon Bull Rev* 1994;1:476–490.
- Macaluso E, Frith CD, Driver J. Modulation of human visual cortex by crossmodal spatial attention. *Science* 2000;289:1206–1208. [PubMed: 10947990]
- Mazza V, Turatto M, Rossi M, Umiltà C. How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia* 2007;45:514–522. [PubMed: 16581094]
- McDonald JJ, Teder-Sälejärvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. *Nature* 2000;407:906–908. [PubMed: 11057669]
- McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA. Neural substrates of perceptual enhancement by cross-modal spatial attention. *J Cogn Neurosci* 2003;15:10–19. [PubMed: 12590839]
- McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA. Neural basis of auditory-induced shifts in visual time-order perception. *Nat Neurosci* 2005;8:1197–1202. [PubMed: 16056224]
- Meredith MA, Stein BE. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 1986;365:350–354. [PubMed: 3947999]
- Middlebrooks JC, Green DM. Sound localization by human listeners. *Annu Rev Psychol* 1991;42:135–159. [PubMed: 2018391]
- Poirier C, Collignon O, Scheiber C, Renier L, Vanlierde A, Tranduy D, Veraart C, De Volder AG. Auditory motion perception activates visual motion areas in early blind subjects. *Neuroimage* 2006;31:279–285. [PubMed: 16443376]
- Rinne T, Anttila S, Winkler I. Mismatch negativity is unaffected by top-down predictive information. *Neuroreport* 2001;12:2209–2213. [PubMed: 11447336]
- Röder B, Stock O, Bien S, Neville H, Rösler F. Speech processing activates visual cortex in congenitally blind humans. *Eur J Neurosci* 2002;16:930–936. [PubMed: 12372029]
- Rozas C, Frank H, Heynen AJ, Morales B, Bear MF, Kirkwood A. Developmental inhibitory gate controls the relay of activity to the superficial layers of the visual cortex. *J Neurosci* 2001;21:6791–6801. [PubMed: 11517267]
- Schoups AA, Vogels R, Orban GA. Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularly. *J Physiol* 1995;483:797–810. [PubMed: 7776259]
- Schoups A, Vogels R, Qian N, Orban G. Practising orientation identification improves orientation coding in V1 neurons. *Nature* 2001;412:549–553. [PubMed: 11484056]
- Schwartz S, Maquet P, Frith C. Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proc Natl Acad Sci USA* 2002;99:17137–17142. [PubMed: 12446842]
- Seitz A, Watanabe T. Is subliminal learning really passive? *Nature* 2003;422:36. [PubMed: 12621425]
- Seitz A, Kim R, Shams L. Sound facilitates visual learning. *Curr Biol* 2006;16:1422–1427. [PubMed: 16860741]
- Smith AT, Singh KD, Williams AL, Greenlee MW. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb Cortex* 2001;11:1182–1190. [PubMed: 11709489]
- Sowden PT, Rose D, Davies IRL. Perceptual learning of luminance contrast detection: specific for spatial frequency and retinal location but not orientation. *Vis Res* 2002;42:1249–1258. [PubMed: 12044757]
- Spence C, Driver J. Audiovisual links in endogenous covert spatial attention. *J Exp Psychol Hum Percept Perform* 1996;22:1005–1030. [PubMed: 8756965]
- Spence C, Driver J. Audiovisual links in exogenous covert spatial orienting. *Percept Psychophys* 1997;59:1–22. [PubMed: 9038403]
- Strybel TZ, Fujimoto K. Minimum audible angles in the horizontal and vertical planes: effects of stimulus onset asynchrony and burst duration. *J Acoust Soc Am* 2000;108:3092–3095. [PubMed: 11144604]

- Swets JA. The relative operating characteristic in psychology. *Science* 1973;182:990–1000. [PubMed: 17833780]
- Teder-Sälejärvi WA, Münte TF, Sperlich F-J, Hillyard SA. Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Cogn Brain Res* 1999;8:327–343.
- Tsushima Y, Sasaki Y, Watanabe T. Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* 2006;314:1786–1788. [PubMed: 17170308]
- Watanabe T, Náñez JE, Sasaki Y. Perceptual learning without perception. *Nature* 2001;413:844–848. [PubMed: 11677607]
- Watanabe T, Náñez JE, 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. *Nat Neurosci* 2002;5:1003–1009. [PubMed: 12219093]
- Weeks R, Horwitz B, Aziz-Sultan A, Tian B, Wessinger CM, Cohen LG, Hallett M, Rauschecker JP. A positron emission tomographic study of auditory localization in the congenitally blind. *J Neurosci* 2000;20:2664–2672. [PubMed: 10729347]
- Xiao DK, Raiguel S, Marcar V, Orban GA. The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cereb Cortex* 1997;7:662–677. [PubMed: 9373021]

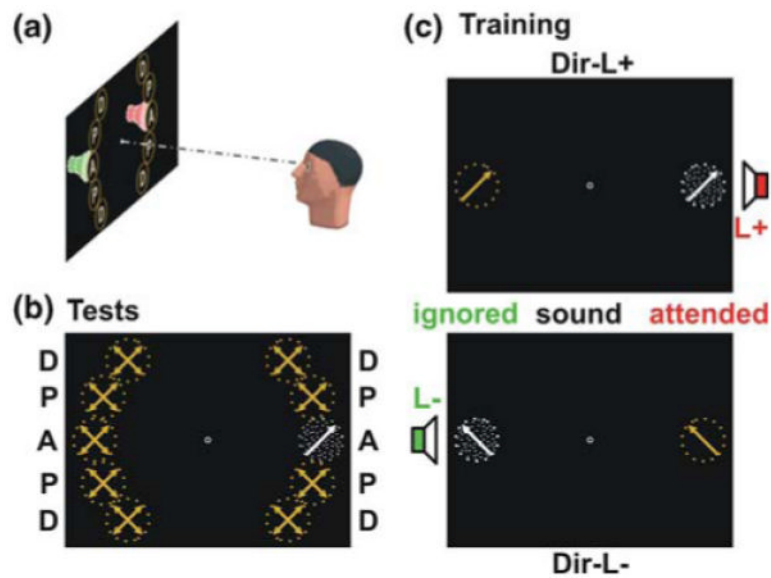


Fig. 1. Experimental paradigm. **a** Experimental setup. Ten locations (left vs. right, A aligned, P proximal, D distal) were tested. Virtual sound sources overlapped with moving dots at location A. **b** During visual test sessions (pre- and post-training) direction discrimination for moving dots (4 directions) was tested for each location separately. **c** During training, participants had to detect deviant sounds at one side (L+). One motion direction (Dir-L+), presented either to the left or right, was paired with this target sound. A perpendicular direction (Dir-L-) was paired with the same sound at the ignored side (L-)

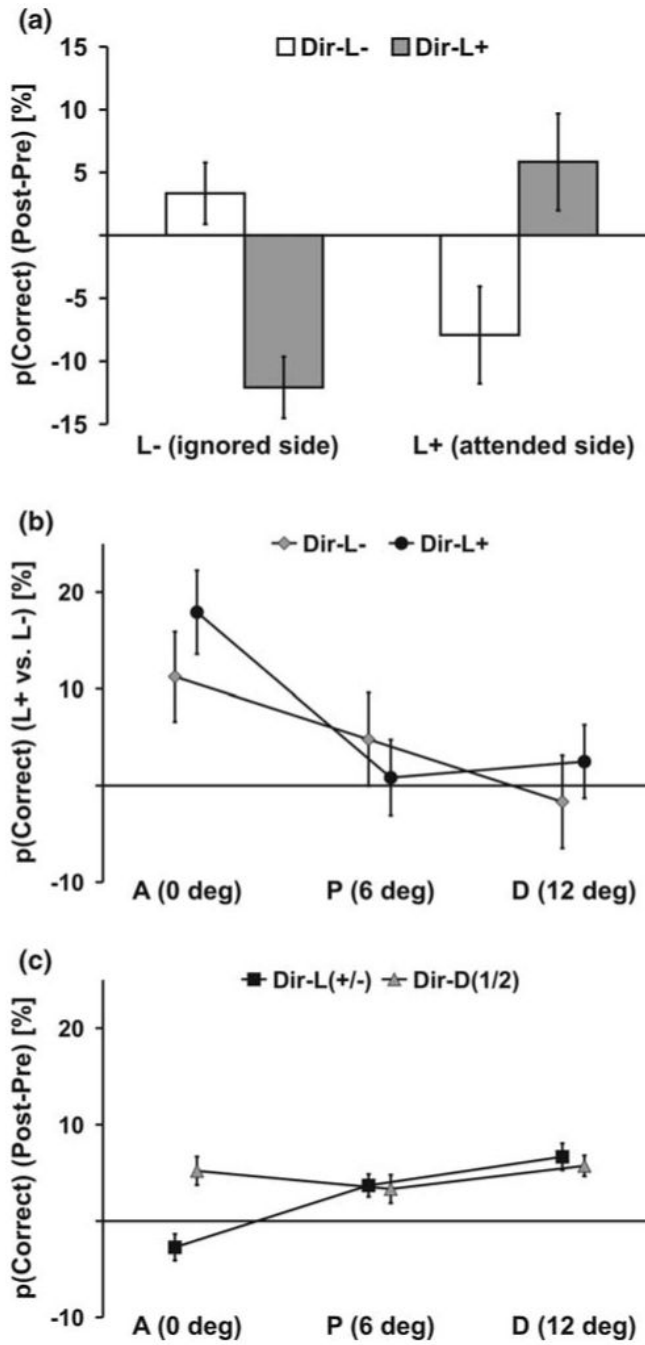


Fig. 2. PL effect (difference percent correct detections post- minus pre-test) in motion discrimination. **a** Performance improved for Dir-L+ at the side where relevant sounds were presented during training (L+) relative to the opposite side. The learning effect was stronger at the irrelevant location (L-) for Dir-L-. **b** This crossmodal PL effect (difference between L+ and L-) was limited to aligned locations (A). **c** Unimodal learning effects for distractor directions (Dir-D1/2) were observed at all visual field locations. Learning effects for trained directions (Dir-L+/-) were reduced compared with learning effects of distractor directions (at A only). *Error bars* reflect within-subjects SEM (Loftus and Masson 1994)

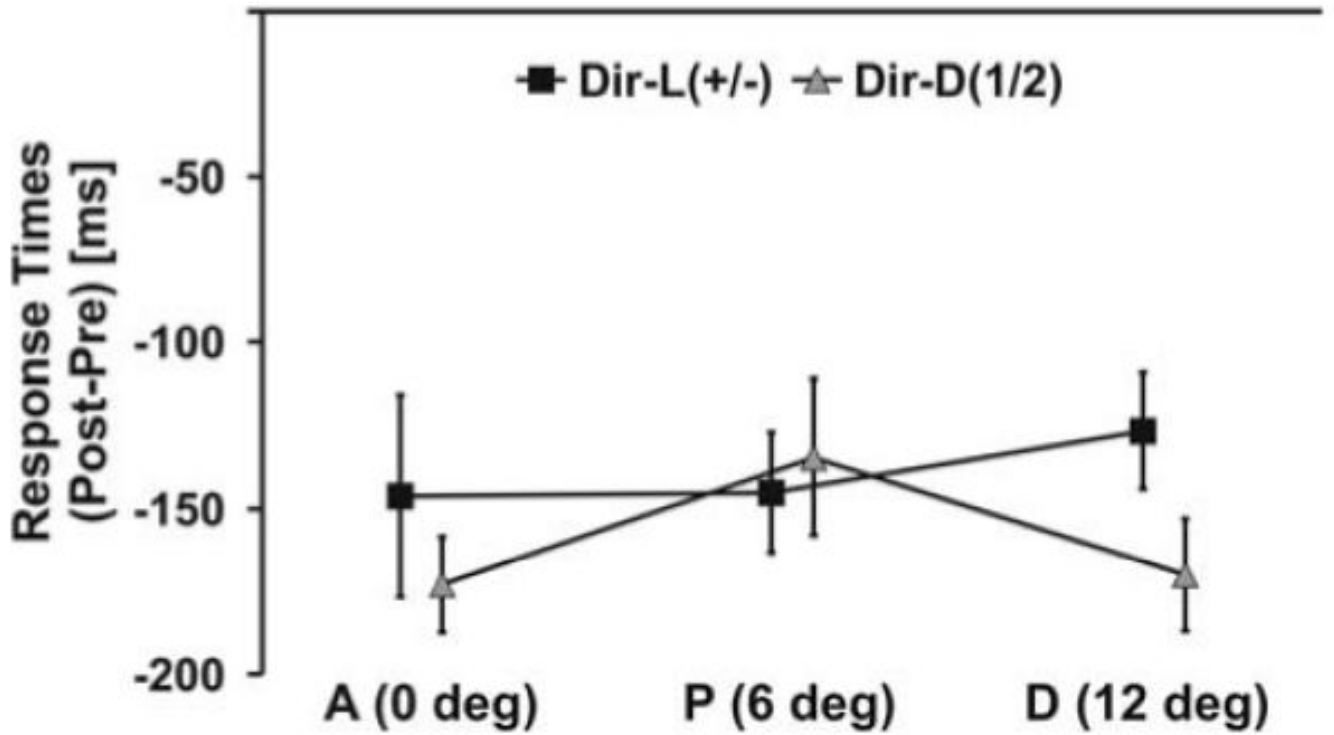


Fig. 3. Learning effect (difference response times post- minus pre-test) for motion discrimination. Observers responded faster at the post-test than at the pre-test irrespective of the tested aperture location (*A* aligned, *P* proximal, *D* distal) and the tested motion direction (Dir-D1/2 distractor direction, Dir-L+/- trained directions). *Error bars* reflect within-subjects SEM (Loftus and Masson 1994)

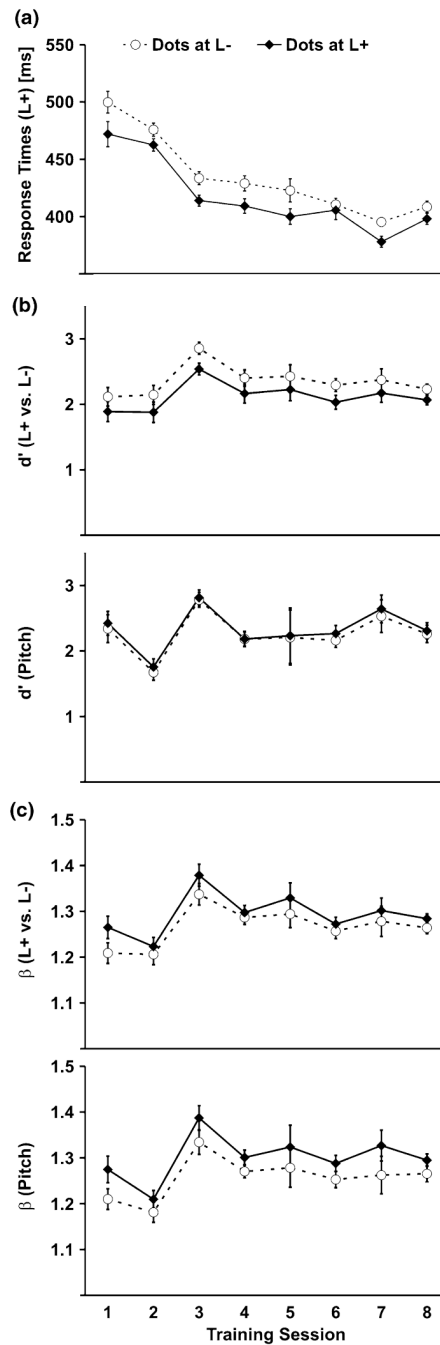


Fig. 4. Auditory training performance for sound detection. **a** Response times to target sounds presented at the relevant location (L+) decreased across training sessions. **b** Sensitivity (d') for discriminating the sound location (L+ vs. L-) was lower when dots appeared at the same location as the sound (dots at L+) than at the opposite side (dots at L-). The location of the dots had no effect on measures of d' for discriminating the pitch of the sounds. **c** Participants were more likely to incorrectly respond to distractor sounds when dots appeared at the same location as the sound (dots at L+) as indicated by a positive shift in response bias (β). Error bars reflect within-subjects SEM (Loftus and Masson 1994)