

Sound Facilitates Visual Learning

Aaron R. Seitz,^{1,3,*} Robyn Kim,^{2,3} and Ladan Shams²

¹Department of Psychology
Boston University
64 Cummings St.

Boston, Massachusetts 02215

²Department of Psychology
University of California, Los Angeles
Los Angeles, California 90095

Summary

Numerous studies show that practice can result in performance improvements on low-level visual perceptual tasks [1–5]. However, such learning is characteristically difficult and slow, requiring many days of training [6–8]. Here, we show that a multisensory audiovisual training procedure facilitates visual learning and results in significantly faster learning than unisensory visual training. We trained one group of subjects with an audiovisual motion-detection task and a second group with a visual motion-detection task, and compared performance on trials containing only visual signals across ten days of training. Whereas observers in both groups showed improvements of visual sensitivity with training, subjects trained with multisensory stimuli showed significantly more learning both within and across training sessions. These benefits of multisensory training are particularly surprising given that the learning of visual motion stimuli is generally thought to be mediated by low-level visual brain areas [6, 9, 10]. Although crossmodal interactions are ubiquitous in human perceptual processing [11–13], the contribution of crossmodal information to perceptual learning has not been studied previously. Our results show that multisensory interactions can be exploited to yield more efficient learning of sensory information and suggest that multisensory training programs would be most effective for the acquisition of new skills.

Results and Discussion

We live in a world in which all of our senses are constantly engaged. When we eat, we appreciate the visual appearance of the food at the same time as the aromas and tastes interact to produce a complete culinary experience. When we converse, we use both sights and sounds to judge the meanings that others express. As we navigate the world, we use sights, sounds, and touch to keep our balance, stay on path, and avoid obstacles.

In the scientific community, perception has traditionally been viewed as a modular function with the different sensory modalities operating largely as separate and

independent systems. However, recently accumulating reports of crossmodal interactions in various perceptual tasks and settings suggest that interactions between modalities are the rule rather than the exception in human processing of sensory information [11–13]. Furthermore, recent neurophysiological studies of crossmodal interactions [14–22] provide accumulating evidence for crossmodal interactions occurring at early stages of perceptual processing in brain areas that have long been viewed as modality specific.

Considering that crossmodal interactions are ubiquitous in human perceptual processing, and that they play some role in the development of perceptual processes [13, 23–25], it is not unlikely that multisensory interactions may be exploited to render the processing of sensory information more effective in terms of encoding and learning as well. To test this idea, we examined the possible benefits of auditory stimuli on low-level perceptual learning of visual-motion, i.e., improvements of motion sensitivity that require extensive training. Low-level perceptual learning is a good test bed for multisensory benefits, because learning is typically very slow, requiring many days of training [6, 7, 26–28], and has been shown to be mediated by early visual areas of the brain [10, 26, 29–31] that are traditionally thought to be highly unimodal [23]. We chose an auditory-visual task because visual motion stimuli are typically accompanied by sounds and because there exists anatomical evidence in animals [32, 33] and a wealth of human psychophysical [11, 13, 34] and neurophysiological [18, 19, 21, 22, 35] evidence indicating vigorous auditory-visual interactions that influence visual processing even in primary visual cortex [36].

We trained two groups of subjects for 10 days each with multisensory stimuli for one group and unisensory stimuli for the other group. The multisensory training consisted of a novel auditory-visual motion-detection task in which both auditory and visual motion stimuli can be used to generate a correct response (see Figure 1): On one-third of trials, only visual signals were present, on one-third of trials only auditory signals were present, and on the remaining one-third, both auditory and visual stimuli were present and were in agreement (see Experimental Procedures). The subjects' task was to detect and discriminate the auditory-visual motion stimuli. Unisensory training consisted of the same task, but with trials containing only visual signals. The number of trials with visual signals was matched between the groups. In this paper, we restrict our analysis to the subset of trials that contained only visual signals for both groups.

Auditory and visual stimuli were made as analogous as possible. Auditory stimuli consisted of Gaussian white noise that produced a percept of a stimulus moving left or right, masked in varying levels of white noise (bandwidth 2–10 KHz, butterworth filtered, ramped). Visual stimuli were dynamic dot patterns of varying motion coherence moving in one of two directions [37]. For each

*Correspondence: aseitz@bu.edu

³These authors contributed equally to this work.

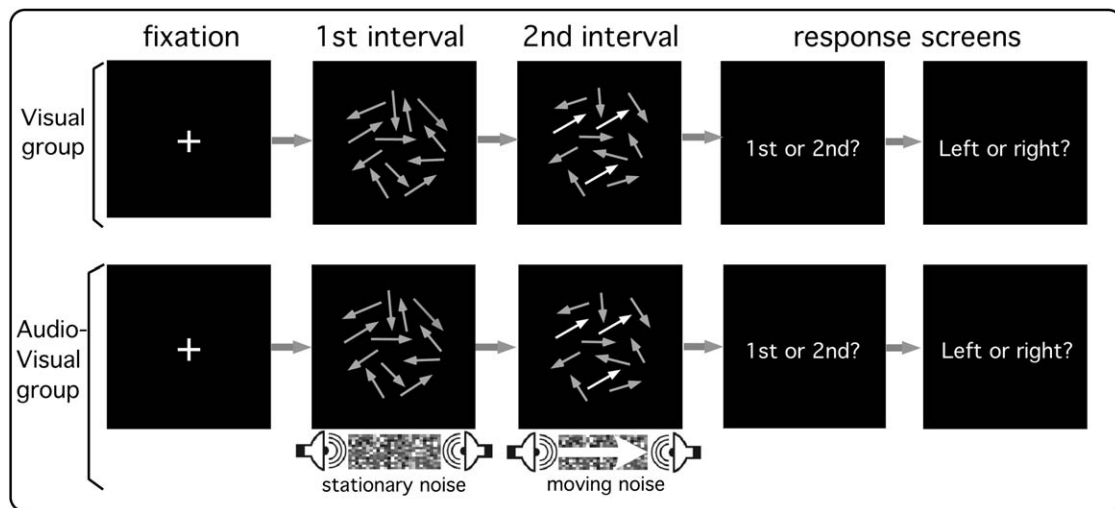


Figure 1. Task Schematic

All sessions for all groups involved the same basic task. A sequence of two stimulus displays was presented (one containing a directional signal and the other containing only noise). After the sequence, subjects reported, with a key press, first the interval that contained the signal (1 or 2) and second the direction of that signal (left or right). In the unisensory condition (top), stimuli consisted of randomly moving dots, a fraction of which moved coherently (direction signal depicted here by white arrows). In the multisensory condition (bottom), visual stimuli were the same and auditory motion stimuli (see [Experimental Procedures](#)) were added, which moved in a direction consistent with that of the visual stimuli. Subjects were required to fixate on a central cross while performing the task. In training sessions, feedback was given on interval response; in tests, no feedback was given.

subject, three different signal levels were selected for each of the visual and auditory modalities through practice sessions, which were run on days previous to that of the training (see [Experimental Procedures](#)).

We first examined performance of subjects in each group during the first training session. Many researchers have reported “fast-learning” effects, whereby perceptual learning has an early phase in which learning will develop within a single session [4, 5, 30, 38, 39]. We find evidence of fast learning in both the multisensory and unisensory groups (see [Figure 2](#)). More importantly, we find that learning occurred significantly more quickly for the multisensory group ([Figure 2](#), black line) than for unisensory group ([Figure 2](#), gray line). A significant learning effect for the multisensory group ($p < 0.01$ paired running t test; black line, [Figure 2](#), bottom panel) is detectable by the second third of the session for the vision-only trials, and was significantly different from that of the unisensory group (see [Figure S1](#) in the [Supplemental Data](#) available online). A bootstrap analysis [40] (see [Experimental Procedures](#)) was run to correct for multiple comparisons and showed a probability of 0.037 of finding more than 15 sequential points below a 5% significance cut-off [41] (as was the case in the running t test for the multisensory group). On the other hand, the unisensory group only showed a tendency of learning, which was delayed until the last third of the session ($p < 0.05$ paired running t test; gray line, [Figure 2](#), bottom panel). Both groups showed deterioration of performance toward the end the session, which has been also reported in other studies of perceptual learning [42]. These results demonstrate that within a single session, multisensory training offers benefits over unisensory training.

It is important to confirm that these learning effects are maintained across sessions. In fact, perceptual

learning of motion often requires multiple days of training to manifest [6, 7, 10, 26], and in some cases the manifestation of learning effects have been shown to be sleep dependant [43]. In order to examine these “slow-learning” effects in a manner that would be minimally contaminated by the fast learning, we compared performance across days for only the first third of the trials in each session (we chose the first third of the session because during this phase neither group showed learning and performance was similar across groups in the first session, but results are qualitatively similar for other early phases of the sessions). Subjects in the multisensory group ([Figure 3](#), black) showed highly significant learning across sessions for the vision-only trials [$F(9,36) = 6.06$, $p < 0.001$, one-way repeated-measures ANOVA] and reached near-peak performance by the third training session ($p < 0.05$, paired t test versus session 1; see running t test [Figure 3](#), bottom panel). The multisensory group also exhibited auditory learning for the sound-only trials (see [Figure S4](#)). On the other hand, while the unisensory group also showed a significant visual learning effect [$F(9, 36) = 3.96$, $p < 0.01$, one-way repeated-measures ANOVA], their performance did not asymptote until about the seventh training session ($p < 0.05$, paired t test versus session 1; see running t test [Figure 3](#), bottom panel).

Comparison across groups showed a benefit of multisensory training [$F(1,80) = 18.8$, $p < 0.001$, two-way ANOVA] and a significant effect of learning across groups [$F(9,80) = 4.7$, $p < 0.001$] but no interaction [$F(9,80) = 0.56$, $p = 0.82$]. The lack of an interaction can be explained by the fact that performance of the multisensory group is nonlinear (i.e., it rises quickly and then asymptotes), and thus the difference in the amount of learning between groups is captured by the main effect of performance across groups. Polynomial regression analysis

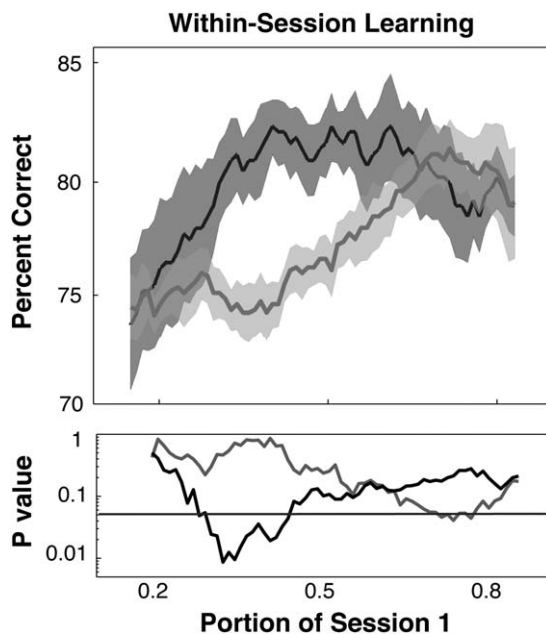


Figure 2. Fast-Learning Results

(Top) Data from first training session for multisensory group (black) and unisensory group (gray) for visual-only trials. Only the detection data are shown here. The discrimination data also show improved learning for the multisensory group (see Figure S2). Ordinate is percent correct averaged across three signal levels of visual-only trials, and abscissa reflects proportion of the first session. Curves were constructed by averaging sequential trials of each signal level for each subject and then smoothing this with a 30-trial-wide boxcar filter (other filter widths provide qualitatively similar results) and averaging smoothed data across subjects; data at the very beginning and end of the session were excluded because of filtering boundary effects. Error bars reflect within-group standard error [46]. (Bottom) Running paired t test for the multisensory (black) and unisensory (gray) groups comparing performance at the beginning of the session (first valid smoothed bin) with each subsequent point in the session on visual-only trials.

confirms that the rate of learning between the two groups is different, with the multisensory group fitting a quadratic function [$F(1,47) = 7.9, p < 0.01$] in contrast to linear learning in the unisensory group [$F(1,48) = 13.5, p < 0.001$].

Although we have demonstrated that the multisensory group learned more effectively than the unisensory group both within and across training sessions in the “unimodal” trials, it is important to note that the unimodal trials for the multisensory group contained a sound stimulus, although it had no directional signal. We thus ran an additional test both to verify that sensitivity improvements in the multisensory trials would transfer to task conditions containing no sounds and also to verify that the training effects were directionally specific (as is expected in perceptual learning). In this test, the sensitivity of subjects in both groups was tested on six different directions of motion (equally spaced around the circle) in silence as well as in presence of stationary noise (see Experimental Procedures for details).

In this test, performance within the multisensory group did not differ between the silent and noise conditions [$F(1,24) = 0.008, p = 0.93$, two-way ANOVA, sound times direction], indicating that the stationary noise

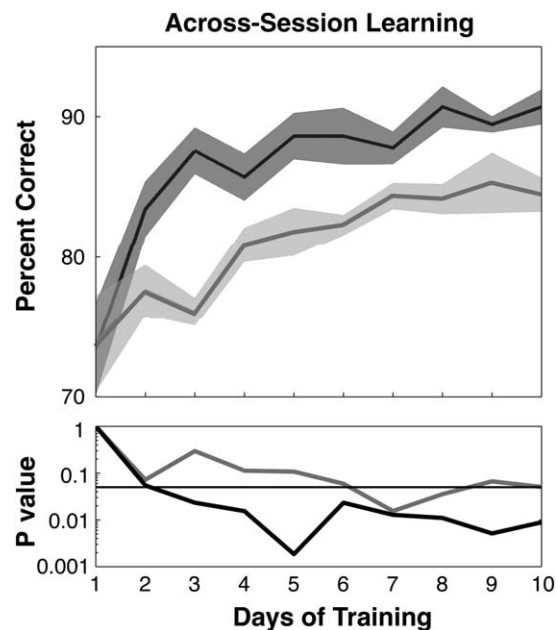


Figure 3. Slow-Learning Results

(Top) Data from each training session for multisensory group (black) and unisensory group (gray) on visual-only trials. Only the detection data are shown here. The discrimination data are qualitatively similar (see Figure S3), as is the effect when broken down by individual stimulus level (see Figure S5). Ordinate is percent correct averaged across three signal levels of visual-only trials during the first third of the session; abscissa reflects training session number. We chose the first third of the session because during this phase neither group showed fast learning and performance was similar across groups in the first session, but results are qualitatively similar for other early phases of the sessions. Error bars reflect within-group standard error [46]. (Bottom) Running paired t test for the multisensory (black) and unisensory (gray) groups comparing performance of the first training session with that of each subsequent session in visual-only trials.

does not provide an enhanced performance compared to no sound. We also found significant effect of direction [$F(5,24) = 3.9, p < 0.01$] with no interaction between the sound condition and direction [$F(5,24) = 0.56, p = 0.73$]. Subjects in the multisensory group performed significantly better on the trained versus the untrained directions [$p < 0.01$; paired t test], indicating that the observed improvement in performance reflects low-level perceptual learning [6]. The unisensory group also showed no effect of sound [$F(1,24) = 0.052, p = 0.83$] and failed to show an effect of direction [$F(5,24) = 1.3, p = 0.31$] or any interactions between sound condition and direction [$F(5,24) = 1.2, p = 0.34$]. A t test between the unisensory and multisensory groups failed to show any significant difference in performance of the trained directions across groups ($p = 0.31$), but this was not unexpected given that both groups had reached asymptote by the last training session (even though learning did not reach significance for the unisensory group). Importantly, the direction test confirmed that the learning effects were directionally specific.

To verify that this was not a native directional bias, or one that developed in the practice sessions (see Experimental Procedures), we recruited five new subjects to conduct the practice and test sessions without any

training. Control subjects also showed no difference between the silent and noise conditions [$F(1,24) = 2.0$, $p = 0.23$]. There was no interaction between the two factors [$F(5,24) = 0.55$, $p = 0.73$] and no main effect of direction [$F(5,24) = 2.3$, $p = 0.074$], and a paired comparison between trained and untrained directions confirmed that there was no innate performance advantage for the directions used in the training procedures ($p = 0.30$, paired t test). Comparison of performance between the control group and the multisensory groups demonstrated similar performance levels with the untrained directions ($p = 0.26$, t test), but the multisensory group performed better than did the control subjects on the trained directions ($p < 0.05$, t test).

A candidate explanation for the multisensory group's advantage is that presentation of sound has an alerting effect, which could benefit the multisensory subjects' task performance and lead to greater learning. We consider this explanation to be unlikely for two reasons. First, in the initial training session, both groups begin training with nearly identical performance, with their paths diverging later in the session; if sound has an alerting effect, it would cause an advantage from the beginning. Second, there is no difference between a silent and a noise condition (in the direction test) after training. Neither one of these by itself is conclusive evidence against alerting, but the two put together (no difference at the beginning or at the end) strongly suggest against it.

Considering our results, we propose that training with an audio-visual stimulus produces enhanced learning both within a session (i.e., fast learning) and across days, most likely by facilitating encoding and promoting better consolidation or retention of learning. Although our results suggest that with time the unisensory training procedure may be able to yield similar performance levels as those produced with multisensory training, the multisensory training is far more efficient. This finding presents the exciting prospect that multisensory training can reduce the number of training days, and perhaps the length of training in each day, needed to produce equivalent performance on a visual perceptual task.

Conclusions

The benefit of sound found in this study is particularly surprising because the task produces a directionally specific learning effect, which is typically thought to be mediated by low-level, purely visual processes. Although it is perhaps to be expected that practice on an audio-visual task should increase audio-visual interactions, it is intriguing that audio-visual practice contributes to enhanced performance on unimodal trials. However, these results are consistent with results showing that auditory motion modulates visual area MT/V5 [44], the visual area most highly specialized for visual motion processing. Additionally, our results are consistent with studies showing enhanced visual recognition of stimuli that were initially presented audio-visually [45]. Notably, semantic congruency was required for enhanced visual recognition. In our experiment, audio-visual directions were congruent; it remains to be explored whether congruency is required, or whether any type of sound can aid learning.

At a broader level, our findings are consistent with a general paradigm of sensory processing in which

perceptual and cognitive mechanisms are tuned to process multisensory signals. Under such a regime, encoding, storing, and retrieving perceptual information is intended by default to operate in a multisensory environment, and unisensory processing is suboptimal because it would correspond to an artificial mode of processing that does not utilize the perceptual machinery to its fullest potential.

Experimental Procedures

Subjects

Fifteen paid subjects (aged 19–24) with normal or corrected-to-normal vision and normal hearing were recruited from the university population and randomly assigned to multisensory ($n = 5$), unisensory ($n = 5$), or direction-control ($n = 5$) groups.

Visual Stimuli

Visual stimuli were dynamic dot patterns of low motion coherence moving in one of two directions, presented at four levels of coherence, including zero coherence and three others tailored to each subject (see *Stimulus Levels*, below). A Newsome-type motion algorithm [37] was employed with white dots (0.2 degree radius) in a 1° – 10° annulus with a dot density of 16.7 dots per deg^2/s and dot speed of 12 deg/s . In this motion algorithm, the subset of coherently moving dots is newly chosen in each frame, and the probability of a given dot lasting more than one frame is the same as the coherence level. Because perception of cardinal directions may be robust to training, we chose 10° (instead of 0°) for rightward motion and 190° (instead of 180°) for leftward motion.

Auditory Stimuli

We designed auditory motion stimuli to be analogous to the visual motion stimuli. Auditory stimuli consisted of Gaussian white noise that produced a percept of a stimulus moving left or right, masked in varying levels of white noise (bandwidth 2–10 KHz, butterworth filtered, ramped). Auditory motion was created by varying amplitude linearly (70–76 dB) between left and right speakers over 300 ms. Speakers were placed on the left and right side of the monitor with the midpoint between the speakers aligned with the fixation point. This produced the perception of sounds that were colocalized with the visual motion stimulus. Different signal levels were created by varying the amplitude ratio between the auditory motion and the noise mask (i.e., computing a weighted average of auditory-motion signal and noise mask). Stimuli at all signal-to-noise levels were normalized so that each had the same root mean square and produced a reading of 76 dB on a sound-pressure meter.

Stimulus Levels

Visual and auditory levels for each subject were determined by plotting psychometric functions for each modality alone on the day before the pretraining tests. For each subject, we chose levels that approximately corresponded to 55%–60%, 70%–75%, and 80%–85% correct detection. The mean coherence levels for each group did not significantly differ (multisensory [low 6.6 ± 0.4 , mid 11 ± 1 , high 15.4 ± 1.8]; unisensory [low 6.8 ± 0.8 , mid 10.4 ± 1.5 , high 14 ± 2.2]).

Task

The task consisted of a two-interval forced-choice task in which observers first reported in which interval (first or second) they detected a directional stimulus and second reported in which direction they discriminated the stimulus to be moving (left or right) by pressing keys on the keyboard.

Procedure

The experiment took place over 14 days. All elements of the experiment were the same for all groups except for training (session numbers 3–12). The first two days served primarily to acclimate the subjects to the task (and thus minimize task-learning effects during training) and determine appropriate *stimulus levels*. These practice tests included unimodal stimuli (audio and visual) at easy coherence levels. Training sessions were conducted for 10 days and included

feedback for the detection response, but not for the discrimination response.

The experimental design was constructed to match the number of trials containing visual signals across the groups. Thus, the unisensory training consisted of three visual levels times two directions with 96 trials of each for a total of 676 trials with visual signals; there were an additional 50 trials containing no signal. Multisensory training had three visual levels times two directions with 48 trials of each, three visual levels times 3 auditory levels times two directions with 16 trials of each, and three auditory levels times two directions with 48 trials of each. This yielded a total of 288 visual-only trials, 288 audiovisual trials, and 288 auditory-only trials; there were also 80 no-signal trials. All trial types were randomly interleaved. For the multisensory group, trials with no auditory directional signal were considered "unisensory" trials. For trials containing both auditory and visual directional signals, the directions were always congruent across modalities; however, the coherence levels were not correlated. Whereas both groups saw the same number of visual stimuli, the multisensory group received additional auditory-alone trials. The direction control group did not conduct the training sessions.

Direction Test Session

In the test session, subjects conducted the same basic task as in the other sessions, but were tested on six different directions of visual motion (10, 70, 130, 190, 250, and 310 degrees of visual angle). In addition, in some trials, auditory noise was presented, and in other trials, no sound (silence) was presented in both intervals. In this way, we could examine whether the learning effects were specific to the trained directions and whether learning transferred between the noise and silent conditions.

Bootstrap Analysis

To validate the significance of the within-session learning effects, we ran a Monte Carlo simulation of our analysis [40]. To do this, we took the average performance for each level for each subject and created a simulated session preserving these averages (i.e., no learning). We then averaged these data across levels and subjects, smoothed them with a boxcar filter, and calculated a running t test in the same manner as was done for the data presented in Figure 2. This procedure was repeated 10,000 times to generate a probability distribution of what result would be expected from chance. To test significance of the experimental data, we calculated the proportion, which turned out to be $p = 0.037$, of simulated sessions in which 15 sequential samples within the session fell below a 5% significance cut-off [41].

Supplemental Data

Supplemental Data include five figures and are available with this article online at: <http://www.current-biology.com/cgi/content/full/16/14/1422/DC1>.

Acknowledgments

We thank Zili Liu for extensive and insightful discussions. We also thank Stephen Engel, Todd Herrington, James Thomas, and Takeo Watanabe for their helpful comments on the manuscript. This project was partly funded by a grant from UCLA academic Senate, and partly by a grant from Naval Research Laboratory (N00173-06-1-G016). R.K. was supported by an National Sciences Foundation fellowship (2005022442), and A.S. was supported by a National Institutes of Health grant (R01EY015980-01).

Received: March 21, 2006

Revised: May 14, 2006

Accepted: May 17, 2006

Published: July 24, 2006

References

1. Adini, Y., Sagi, D., and Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature* 415, 790–793.

- Seitz, A.R., Nanez, J.E., Holloway, S.R., and Watanabe, T. (2005). Visual experience can substantially alter critical flicker fusion thresholds. *Hum. Psychopharmacol.* 20, 55–60.
- Seitz, A.R., Yamagishi, N., Werner, B., Goda, N., Kawato, M., and Watanabe, T. (2005). Task-specific disruption of perceptual learning. *Proc. Natl. Acad. Sci. USA* 102, 14895–14900.
- Fiorentini, A., and Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature* 287, 43–44.
- Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science* 256, 1018–1021.
- Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature* 387, 401–406.
- Ball, K., and Sekuler, R. (1981). Adaptive processing of visual motion. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 780–794.
- Seitz, A., and Watanabe, T. (2005). A unified model for perceptual learning. *Trends Cogn. Sci.* 9, 329–334.
- Fahle, M., Edelman, S., and Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Res.* 35, 3003–3013.
- Watanabe, T., Nanez, J.E., Sr., Koyama, S., Mukai, I., Liederman, J., and Sasaki, Y. (2002). Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nat. Neurosci.* 5, 1003–1009.
- Calvert, G., Spence, C., and Stein, B.E. (2004). *The Handbook of Multisensory Processes* (Cambridge, MA: The MIT Press).
- Driver, J., and Spence, C. (1998). Crossmodal attention. *Curr. Opin. Neurobiol.* 8, 245–253.
- Shimojo, S., and Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Curr. Opin. Neurobiol.* 11, 505–509.
- Zangaladze, A., Epstein, C.M., Grafton, S.T., and Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature* 401, 587–590.
- Macaluso, E., Frith, C.D., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science* 289, 1206–1208.
- Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, O.W., Iversen, S.D., and David, A.S. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596.
- Amedi, A., Malach, R., Hendler, T., Peled, S., and Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330.
- Giard, M.H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490.
- Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., and Foxe, J.J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128.
- Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., and Schroeder, C.E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.* 10, 77–83.
- Shams, L., Kamitani, Y., Thompson, S., and Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport* 12, 3849–3852.
- Bhattacharya, J., Shams, L., and Shimojo, S. (2002). Sound-induced illusory flash perception: Role of gamma band responses. *Neuroreport* 13, 1727–1730.
- Pascual-Leone, A., Amedi, A., Fregni, F., and Merabet, L.B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci.* 28, 377–401.
- Carlsen, R., and Lickliter, R. (1999). Augmented prenatal tactile and vestibular stimulation alters postnatal auditory and visual responsiveness in bobwhite quail chicks. *Dev. Psychobiol.* 35, 215–225.
- Sleigh, M.J., and Lickliter, R. (1998). Timing of presentation of prenatal auditory stimulation alters auditory and visual responsiveness in bobwhite quail chicks (*Colinus virginianus*). *J. Comp. Psychol.* 112, 153–160.

26. Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature* *412*, 549–553.
27. Seitz, A.R., and Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature* *422*, 36.
28. Seitz, A., Lefebvre, C., Watanabe, T., and Jolicoeur, P. (2005). Requirement for high-level processing in subliminal learning. *Curr. Biol.* *15*, R753–R755.
29. Furmanski, C.S., Schluppeck, D., and Engel, S.A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Curr. Biol.* *14*, 573–578.
30. Vaina, L.M., Belliveau, J.W., des Roziers, E.B., and Zeffiro, T.A. (1998). Neural systems underlying learning and representation of global motion. *Proc. Natl. Acad. Sci. USA* *95*, 12657–12662.
31. Zohary, E., Celebrini, S., Britten, K.H., and Newsome, W.T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science* *263*, 1289–1292.
32. Rockland, K.S., and Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Int. J. Psychophysiol.* *50*, 19–26.
33. Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.* *22*, 5749–5759.
34. De Gelder, B., and Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.* *7*, 460–467.
35. Shams, L., Iwaki, S., Chawla, A., and Bhattacharya, J. (2005). Early Modulation of visual cortex by sound: An MEG study. *Neurosci. Lett.* *378*, 76–81.
36. Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., and Rees, G. Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage*, *31*, 1247–1256. Published online March 23, 2006. 10.1016/j.neuroimage.2006.01.016.
37. Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.* *12*, 4745–4765.
38. Chun, M.M., and Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognit. Psychol.* *36*, 28–71.
39. Hawkey, D.J., Amitay, S., and Moore, D.R. (2004). Early and rapid perceptual learning. *Nat. Neurosci.* *7*, 1055–1056.
40. Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap* (New York: Chapman & Hall).
41. Rugg, M.D., Doyle, M.C., and Melan, C. (1993). An event-related potential study of the effects of within- and across-modality word repetition. *Lang. Cogn. Process.* *8*, 357–377.
42. Mednick, S.C., Arman, A.C., and Boynton, G.M. (2005). The time course and specificity of perceptual deterioration. *Proc. Natl. Acad. Sci. USA* *102*, 3881–3885.
43. Mednick, S., Nakayama, K., and Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. *Nat. Neurosci.* *6*, 697–698.
44. Lewis, J.W., Beauchamp, M.S., and DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* *10*, 873–888.
45. Lehmann, S., and Murray, M.M. (2005). The role of multisensory memories in unisensory object discrimination. *Brain Res. Cogn. Brain Res.* *24*, 326–334.
46. Loftus, G.R., and Masson, M.E. (1994). Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* *1*, 476–490.

Supplemental Data

Sound Facilitates Visual Learning

Aaron R. Seitz, Robyn Kim, and Ladan Shams

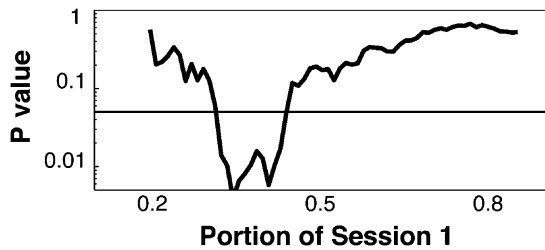


Figure S1. Between-Group Comparison for Visual-Detection Fast Learning

The figure shows a running t test comparing performance between the multisensory and unisensory groups on training day 1. Comparisons are made on data from visual-only trials, and the t test is performed on data after smoothing (see Figure 2 for details of performance data for each group).

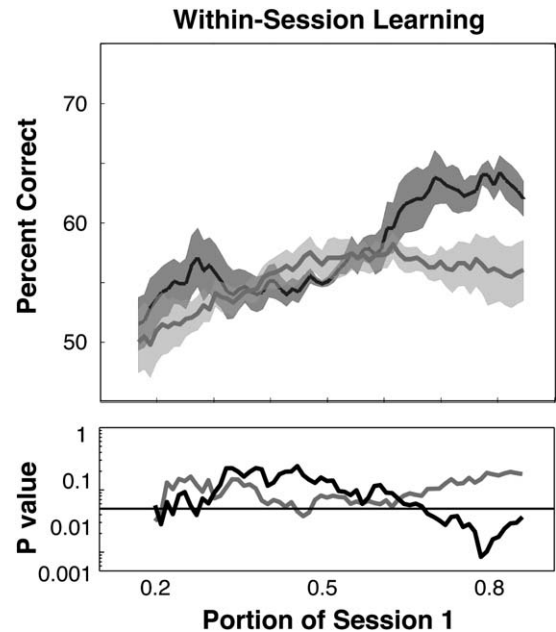


Figure S2. Visual-Discrimination Fast-Learning results

(Top) Data from first training session for the multisensory group (black) and the unisensory group (gray). Ordinate is percent correct averaged across three signal levels of visual-only trials, and abscissa reflects proportion of the first session. Curves were constructed by averaging sequential trials of each signal level for each subject and then smoothing this with a 30-trial-wide boxcar filter (other filter widths provide qualitatively similar results) and averaging smoothed data across subjects; data at the very beginning and end of the session were excluded because of filtering boundary effects. Error bars reflect within-group standard error.

(Bottom) Running paired t test within the multisensory (black) and unisensory (gray) groups, with performance compared at the front of the session (first valid smoothed bin) with each subsequent point in the session.

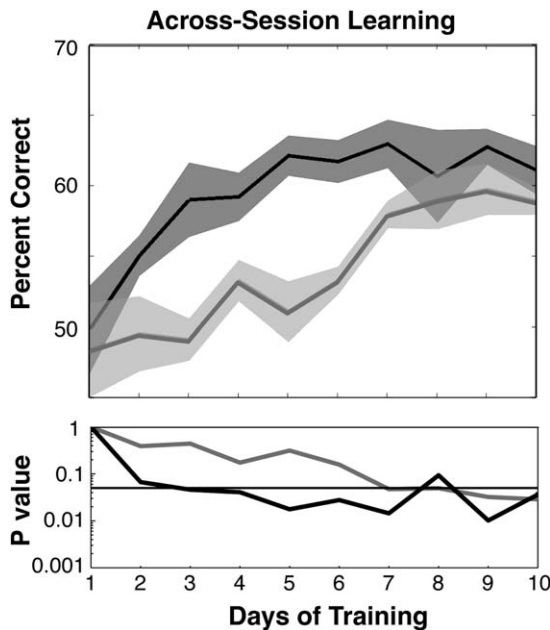


Figure S3. Visual-Discrimination Slow-Learning Results

(Top) Data from each training session for the multisensory group (black) and the unisensory group (gray). Ordinate is percent correct averaged across three signal levels of visual-only trials, and abscissa reflects session number. Data are the average performance on unisensory trials of each group during the first third of each session. Subjects in the multisensory group showed significant learning across sessions [$F(9,36) = 2.81, p < 0.01$, one-way repeated-measures ANOVA] and reached near-peak performance by the fifth training session ($p < 0.05$, see bottom panel) and reached significance on the third day. On the other hand, although the unisensory group also showed a significant learning effect [$F(9,36) = 3.99, p < 0.01$, one-way repeated-measures ANOVA], its performance did not asymptote or reach significance until about the seventh training session ($p < 0.05$, see bottom panel). Error bars reflect within-group standard error.

(Bottom) Running paired t test for each within the multisensory (black) and unisensory (gray) groups, with performance of the first training session compared with that of each subsequent session.

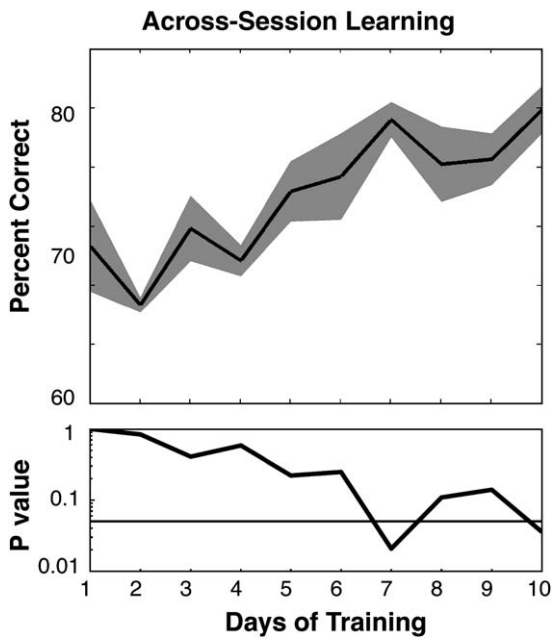


Figure S4. Auditory-Detection Slow-Learning Results

(Top) Data from each training session for the multisensory group; ordinate is percent correct averaged across three signal levels of visual-only trials, and abscissa reflects session number. Data are the average performance on unisensory sound trials during the first third of each session. Error bars reflect within-group standard error. A one-way repeated-measures ANOVA demonstrated a significant learning effect across training sessions [$F(9,36) = 3.2, p < 0.01$]. (Bottom) Running paired t test for comparing performance of the first training session with that of each subsequent session. Note that fast learning was not found for auditory detection.

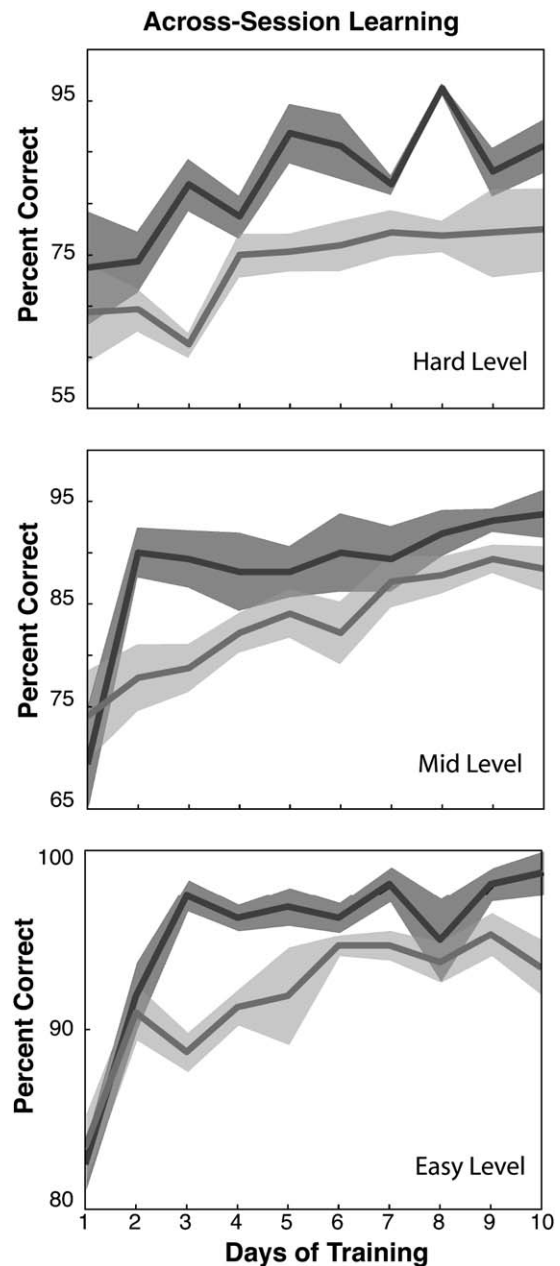


Figure S5. Visual-Detection Slow-Learning Results for Each Difficulty Level

(Top) Data from hardest-difficulty trials in each training session for the multisensory group (black) and the unisensory group (gray). Subjects in the multisensory group showed significant learning across sessions [$F(9,36) = 2.7, p < 0.05$, one-way repeated-measures ANOVA], but subjects in the unisensory group did not [$F(9,36) = 1.4, p = 0.22$].

(Middle) Data from mid-difficulty trials in each training session. Subjects in the multisensory group showed significant learning across sessions [$F(9,36) = 3.6, p < 0.01$, one-way repeated-measures ANOVA], as did subjects in the unisensory group [$F(9,36) = 2.8, p < 0.05$].

(Bottom) Data from easiest-difficulty trials in each training session. Subjects in the multisensory group showed significant learning across sessions [$F(9,36) = 10.0, p < 0.001$, one-way repeated-measures ANOVA], as did subjects in the unisensory group [$F(9,36) = 4.4, p < 0.001$]. Error bars reflect within-group standard error.