

# Hearing Shapes: Event-related Potentials Reveal the Time Course of Auditory–Visual Sensory Substitution

Christian Grauly<sup>1</sup>, Orestis Papaioannou<sup>2</sup>, Phoebe Bauer<sup>1</sup>, Michael A. Pitts<sup>1</sup>,  
and Enriqueta Canseco-Gonzalez<sup>1</sup>

## Abstract

■ In auditory–visual sensory substitution, visual information (e.g., shape) can be extracted through strictly auditory input (e.g., soundscapes). Previous studies have shown that image-to-sound conversions that follow simple rules [such as the Meijer algorithm; Meijer, P. B. L. An experimental system for auditory image representation. *Transactions on Biomedical Engineering*, 39, 111–121, 1992] are highly intuitive and rapidly learned by both blind and sighted individuals. A number of recent fMRI studies have begun to explore the neuroplastic changes that result from sensory substitution training. However, the time course of cross-sensory information transfer in sensory substitution is largely unexplored and may offer insights into the underlying neural mechanisms. In this study, we recorded ERPs to soundscapes before and after sighted participants were trained with the Meijer

algorithm. We compared these posttraining versus pretraining ERP differences with those of a control group who received the same set of 80 auditory/visual stimuli but with arbitrary pairings during training. Our behavioral results confirmed the rapid acquisition of cross-sensory mappings, and the group trained with the Meijer algorithm was able to generalize their learning to novel soundscapes at impressive levels of accuracy. The ERP results revealed an early cross-sensory learning effect (150–210 msec) that was significantly enhanced in the algorithm-trained group compared with the control group as well as a later difference (420–480 msec) that was unique to the algorithm-trained group. These ERP modulations are consistent with previous fMRI results and provide additional insight into the time course of cross-sensory information transfer in sensory substitution. ■

## INTRODUCTION

Sensory substitution occurs when an individual uses one sensory modality to acquire information typically obtained by a different sensory modality. One of the first demonstrations of this phenomenon was in 1969, when Bach-y-Rita and colleagues found that blind individuals were able to access visual information via devices that converted images into tactile sensations delivered to the back (Bach-y-Rita, Collins, Sanders, White, & Scadden, 1969; see also Bach-y-Rita, Kaczmarek, Tyler, & Garcia-Lara, 1998). Sensory substitution has generated a large amount of interest because of its clinical applications, resulting in the development of various sensory substitution devices (SSDs). One such device, known as “The vOICE,” translates visual information into auditory “soundscapes” using a set of image-to-sound conversion rules (Meijer, 1992). In addition to clinical benefits, SSDs have opened new avenues for basic research aimed at identifying the mechanisms that support behavioral and neuroplastic changes associated with cross-modal perceptual learning.

A number of behavioral studies have demonstrated that both blind and sighted participants can perform

high-level visual tasks with strictly auditory input using visual-to-auditory SSDs. For example, trained participants have been successful in discriminating object categories (Cronly-Dillon, Persaud, & Gregory, 1999) as well as individual visual features, such as orientation, size, shape, texture, and relative spatial location (Bermejo, Di Paolo, Hug, & Arias, 2015; Levy-Tzedek, Hanassy, Abboud, Maidenbaum, & Amedi, 2012; Striem-Amit, Guendelman, & Amedi, 2012; Kim & Zatorre, 2008; Arno et al., 2001; Sampaio, Maris, & Bach-y-Rita, 2001; Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart, 1999). Importantly, participants in these studies were able to identify the shapes conveyed by novel soundscapes, ones that had never been presented during training. This successful transfer from trained to novel stimuli demonstrates a generalizable form of perceptual learning based on the image-to-sound conversion rules of the SSD, rather than simple memorization of the auditory–visual stimulus pairs.

The behavioral evidence for successful sensory substitution has led to subsequent fMRI studies investigating the neural mechanisms underlying this phenomenon (Hertz & Amedi, 2015; Campus et al., 2012; Striem-Amit et al., 2012; Kim & Zatorre, 2011; Ortiz et al., 2011; Amedi et al., 2007; Collignon, Lassonde, Lepore, Bastien, & Veraart, 2006; Amedi, Malach, & Pascual-Leone, 2005). For

<sup>1</sup>Reed College, <sup>2</sup>University of California, Davis

example, Amedi et al. (2005, 2007) found that individuals given auditory–visual sensory substitution training show a greater BOLD signal in the lateral occipital complex (LOC) posttraining. This region was previously thought to be strictly involved in visual processing of shape information (Amedi, Malach, Hendler, Peled, & Zohary, 2001; Grill-Spector, Kourtzi, & Kanwisher, 2001; Grill-Spector, Kushnir, Edelman, Itzhak, & Malach, 1998), but it has since been implicated in shape perception independent of sensory modality (Bubic, Striem-Amit, & Amedi, 2010; Amedi et al., 2005, 2007; Peltier et al., 2007; Pietrini et al., 2004). A later study found that sensory substitution training may not necessarily enhance LOC activity posttraining but rather increase functional connectivity between regions of auditory cortex and the LOC (Kim & Zatorre, 2011). This recruitment of nonauditory brain regions as a result of sensory substitution also seems to reach beyond the LOC, including regions within the left precentral sulcus, the right occipito-parietal sulcus, and, to a lesser extent, the right posterior medial occipital cortex (Striem-Amit et al., 2011). Overall, these studies suggest fairly rapid neuroplastic changes due to sensory substitution training.

Although these fMRI studies have demonstrated which regions may be involved in this type of cross-sensory learning, they have not yet been able to elucidate the timing of neural events that support information transfer during sensory substitution. Measuring the time course of the neural processes involved in sensory substitution may help address a number of basic questions. For example, one question commonly raised is whether substitution occurs at an early stage of stimulus processing once individuals are trained. Temporal information derived from ERPs could complement the spatial information found by fMRI research, providing a more complete picture of this unique phenomenon.

A critical challenge for properly addressing such questions is to ensure that the measured brain activity is indeed specific to sensory substitution and does not reflect a more generic form of cross-sensory activation or associative learning. Cross-modal activation in the visual and auditory cortex can be elicited by various paradigms that are independent of sensory substitution (Murray et al., 2016; Feng, Störmer, Martinez, McDonald, & Hillyard, 2014; Grahn, Henry, & McAuley, 2011; Buetti & Macaluso, 2010; Driver & Noesselt, 2007). Indeed, previous studies have noted the inherent difficulty in creating a well-matched control condition, which includes all aspects of the training, including generic cross-sensory activity, but without genuine sensory substitution (Hertz & Amedi, 2015; Campus et al., 2012; Striem-Amit et al., 2012; Kim & Zatorre, 2011; Ortiz et al., 2011; Amedi et al., 2005, 2007; Collignon et al., 2006). To isolate neural events specific to sensory substitution, it is imperative to compare an experimental group trained with SSDs with a control group that can achieve comparable recognition performance on learned pairs of stimuli (but with-

out SSD training), while keeping exposure to stimuli, tasks, and cognitive effort as equivalent as possible.

In this study, we utilized the ERP technique to shed light on the timing of events during auditory–visual sensory substitution. We first recorded ERPs elicited by auditory soundscapes translated from complex visual shapes using “the vOICe” SSD, in naive sighted individuals (pre-training). Half of the participants (hereafter referred to as the “algorithm-trained group”) were then trained to use the SSD, such that they could successfully match soundscapes with their corresponding shapes. To isolate the effects of sensory substitution, the remaining half of participants (the control group) were given the same procedure, with the same stimuli, except that they were trained to memorize randomized shape–soundscape pairs. After training, ERPs elicited by the same soundscapes were recorded again in both groups. We predicted that the algorithm-trained group would exhibit ERP differences (posttraining vs. pre-training) that were larger or unique compared with the control group, potentially revealing the time course of substitution-specific neural activity.

## EXPERIMENT 1

### Methods

#### *Participants*

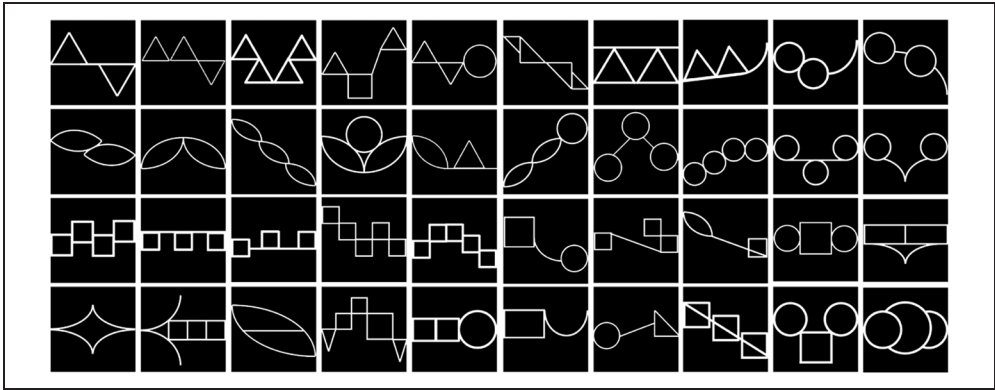
The participants included 36 college students (23 women), aged 19–24 (mean age = 21) years, with normal hearing and normal or corrected-to-normal vision. The participants were split randomly into the algorithm-trained group and the control group. The data of five participants (one algorithm-trained and four control participants) were removed from subsequent analyses, because of greater than 30% of trials per participant being excluded during EEG artifact rejection. The remaining participants included 16 in the algorithm-trained group and 15 in the control group. Each participant attended three sessions, one per day for 3 consecutive days, with each session lasting approximately 2 hr.

Participants were paid \$10 for the first session, \$20 for the second session, and \$30 for the final session. All procedures were approved by the Reed College institutional review board. Written informed consent was obtained from each participant before the first session of the experiment.

#### *Stimuli*

The visual stimuli consisted of 160 complex shapes created from an “alphabet” of 13 basic shapes (see Figure 1 for examples). Shapes were presented as white outlines on a black background. All images were cropped to 215 × 215 pixel squares, with the shapes touching both the right and left edges of the image. All images subtended a viewing angle of 2.4° × 2.4° and were presented using the Presentation software package (Neurobehavioral Systems, Inc., Albany, CA) on a Dell SA2311W monitor with a frame

**Figure 1.** Examples of the visual stimuli used and translated into soundscapes.



rate of 60 Hz. Participants maintained their gaze on a small ( $0.07^\circ$ ) white centrally located fixation dot that was visible throughout each trial.

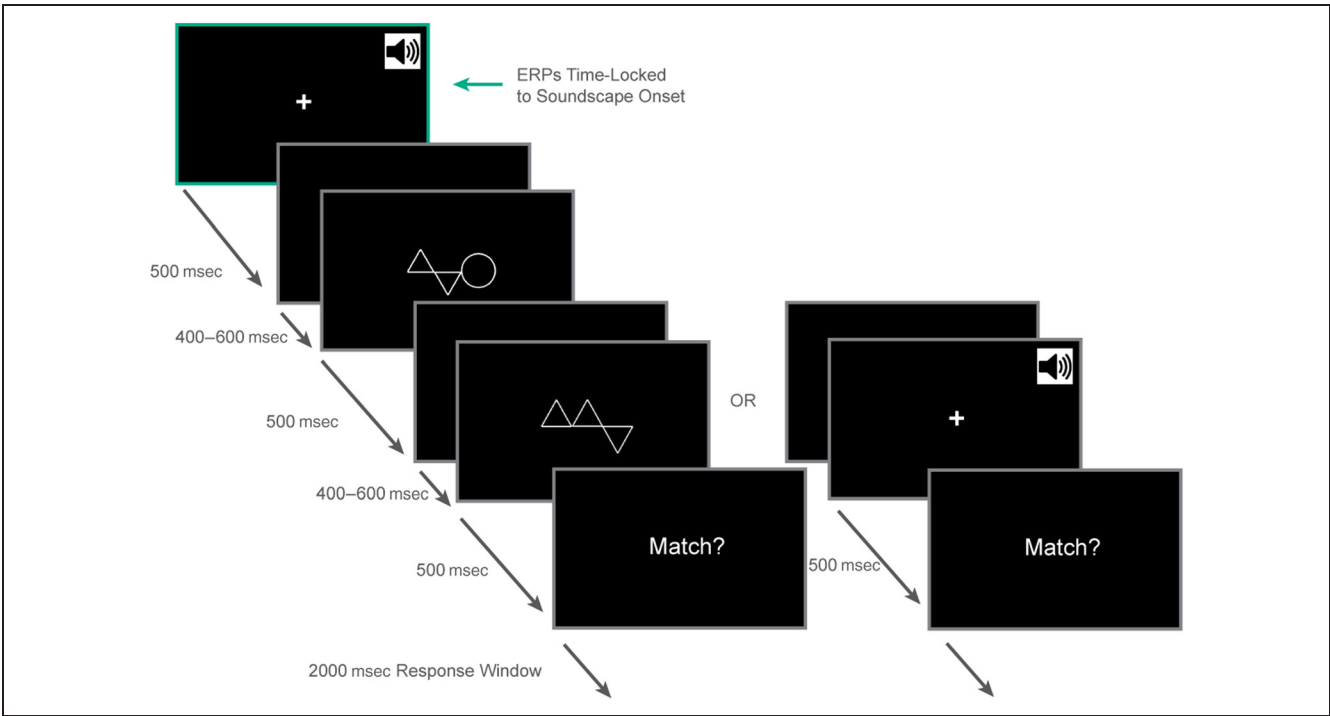
Auditory soundscapes were translated from the 160 visual stimuli using Version 1.93 of the vOICE Learning Edition software for Windows. The vOICE creates sound clips out of images using the Meijer algorithm (Meijer, 1992), which transcribes the vertical axis as pitch (such that higher tones correspond to a higher position in the image) and the horizontal axis as time (such that the beginning of the sound corresponds to the left side of the image and pans from left to right). The Meijer algorithm also translates brightness to loudness, but in the present experiment, brightness was held constant. All soundscapes lasted 500 msec and were presented through Philips SHS4700 earclip headphones. Soundscape stimuli had an average loudness of 65.4 dB, with a range between 59 and 68 dB. Pitch range for these stim-

uli was held at the default levels of the vOICE learning edition, with the lowest possible frequency of 500 Hz and the highest possible frequency of 5000 Hz.

For the algorithm-trained group, each visual stimulus was paired with its corresponding soundscape created via the Meijer conversion algorithm (Meijer, 1992). For the control group, the same visual stimuli were paired with the same set of auditory soundscapes, but the pairings were made randomly (i.e., For this group, each visual stimulus was randomly paired with the soundscape created from another visual stimulus, such that the rules of the Meijer algorithm did not apply to these pairings).

#### *Procedure (Day 1: Pretraining Session with EEG)*

In the pretraining session, participants performed a simple matching task on 80 of the sound–shape pairs while their EEG was recorded (Figure 2). In this task (hereafter



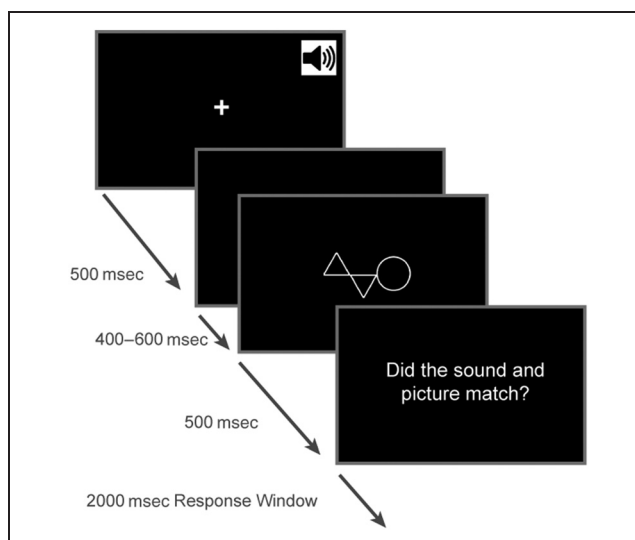
**Figure 2.** Example trial sequence for the triad task used in the pretraining and posttraining sessions.

referred to as the “triad task”), three consecutive stimuli were presented for 500 msec each, separated by a blank/silent ISI of variable duration (400–600 msec). The first stimulus was always a soundscape, the second was always a visual shape, and the third could be either a soundscape (50%) or a visual shape (50%). The third stimulus was identical to either the first (16.7%) or the second (16.7%) or neither (66.7%). After the third stimulus in each triad, participants were given a 2000-msec response window to make a two-alternative forced choice (2-AFC) indicating whether there were any matches or not, that is, whether any two of the three stimuli were identical. Participants used their right index and middle fingers to indicate match and no-match trials, respectively. After each response (or if no response was made within the 2000-msec response window), the next trial began after a variable interval of 900–1100 msec.

Before EEG recording began, participants completed 36 practice trials with soundscapes/shapes that were not part of the main study. The EEG session consisted of 640 total trials, delivered in 10 blocks of 64 trials with a short break after each block. Each block lasted approximately 4 min. A longer break was provided at the halfway point (after five blocks). Each of the 80 soundscapes and visual shapes was presented eight times in the first and second stimulus positions, respectively, and four times in the third stimulus position. The same set of 80 soundscapes and visual shapes was presented to each participant, but the order was randomized across participants. Unbeknownst to the participants and irrelevant to the task in Session 1, on 33.3% of the trials, the second stimulus (visual shape) corresponded to the first stimulus (soundscape) in terms of the Meijer algorithm. The third stimulus (soundscape or shape) never matched the first or second cross-modally (via the Meijer algorithm).

#### *Procedure (Day 2: Training Session, No EEG)*

In the “training” session, participants performed three different training tasks (EEG was not recorded during this session). The algorithm-trained group was informed of the rules that govern the Meijer image-to-sound conversion algorithm and was trained on pairs of soundscapes and visual shapes that followed these rules. The control group was trained on pairs of stimuli that did not follow the rules of the Meijer algorithm and was informed that the pairs were randomly created. The algorithm-trained group was instructed to use the training tasks to learn to “translate” the sounds into images, and the control group was instructed to try to memorize all of the sound–image pairs as best as possible. Both groups performed the same three training tasks. The first was a passive exposure task that presented each of the 80 sound–image pairs simultaneously (in a random order), with each pair being presented three times. The second training was a five-alternative forced choice (5-AFC) task in which a soundscape was presented first, followed by five different visual



**Figure 3.** Two-AFC matching task used as the third training task as well as the transfer test.

shapes presented simultaneously. Participants were asked to select, from the five choices, the correct image associated with the soundscape. Each of the 80 soundscapes was presented four times during the 5-AFC training task. The third training was a 2-AFC matching task in which a soundscape was followed by a visual shape, and participants were asked to determine if they were correctly paired or not. On both the 5-AFC task and the 2-AFC matching task, participants were given feedback after each trial indicating whether they had responded correctly or not.

At the end of Session 2, after completing the three training tasks, participants’ learning was assessed via a transfer test (Figure 3). This transfer test involved a 2-AFC matching task identical to the third training task except that no feedback was provided, and 80 untrained sound–image pairs (stimuli that had not been presented during the training) were mixed in with the 80 trained pairs. The purpose of this test was to assess the extent of generalized perceptual learning in the algorithm-trained group, that is, to test whether these participants were able to learn the rules of the algorithm such that they could extract shape information from novel soundscape stimuli.

#### *Procedure (Day 3: Posttraining Session with EEG)*

In the posttraining session, EEG was recorded while participants performed the same triad task as they did in the pretraining session (Figure 2), except for one change in the instructions. In the posttraining session, participants were asked to indicate whether or not any two of the three stimuli “matched,” including the first (soundscape) and second (visual shape) stimuli. That is, they were instructed to indicate not only unimodal matches (as in Session 1) but also any cross-modal matches (based on their training in Session 2). The stimuli were identical to those presented in the pretraining session, but the

order was randomized in both sessions. Six hundred forty trials were presented, and each trial included a unimodal match (33.3%), a cross-modal match (33.3%), or no match (33.3%).

### EEG/ERP Methods

EEG scalp voltages were recorded using a 96-channel electrode cap with equidistant electrode placements (EasyCap, Herrsching, Germany). Signals were amplified via  $3 \times 32$ -channel BrainAmp Standard amplifiers (Brain Products, Gilching, Germany), online filtered (150-Hz low-pass filter, 0.1-Hz high-pass filter), and digitized at 500 Hz. A 35-Hz low-pass filter (roll-off: 24 dB/oct) was subsequently applied offline. During data collection, all channels were referenced to CPz and were rereferenced to the average of all 92 scalp channels offline. Eye blinks and eye movements were monitored via a vertical EOG channel with an electrode positioned under the left eye and left and right horizontal EOG channels (rereferenced to each other offline to form a bipolar pair) with one electrode positioned lateral to each eye. All sensors were individually adjusted until the impedance was less than 5 k $\Omega$ .

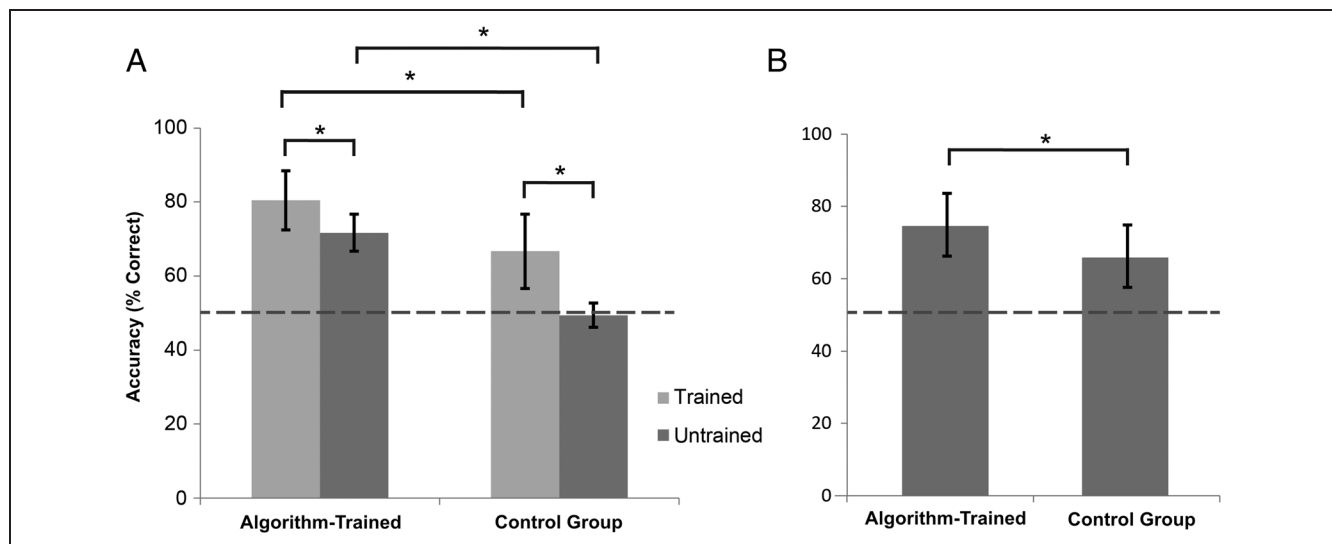
Pretraining and posttraining ERPs were time-locked to the onset of the first stimulus (soundscape) of each triad. Trials with eye movement, blink, or muscle artifacts within a time window of  $-100$  to  $800$  msec (relative to the soundscape onset) were detected and rejected semi-automatically by a combination of computer-based peak-to-peak amplitude thresholds and visual inspection. On average, 18.9% of the 600 trials were rejected because of excessive artifacts (19.2% in the algorithm-trained group, 18.5% in the control group). If more than 30% of trials of a given participant were rejected, that participant's data were excluded from further analysis. Pretraining and post-

training ERPs elicited by the soundscapes in each group were averaged and baseline corrected from  $-100$  to  $0$  msec before soundscape onset.

## Results

### Behavioral Results (Transfer Test)

The transfer test, completed directly after training (Session 2), consisted of a 2-AFC task in which participants had to determine whether a soundscape–shape pair correctly matched based on their training. This test included 50% trained stimuli and 50% untrained (novel) stimuli. Transfer test accuracy (Figure 4A) was evaluated using a mixed-model ANOVA assessing the within-participant factor Stimulus type (trained vs. untrained) and the between-participant factor Group (algorithm-trained vs. control). This ANOVA revealed main effects of Group,  $F(1, 29) = 76.8, p < .0001, \eta^2 = .49$ , and Stimulus type,  $F(1, 29) = 57.5, p < .0001, \eta^2 = .26$ , as well as an interaction between Group and Stimulus type,  $F(1, 29) = 6.34, p = .02, \eta^2 = .32$ . In the algorithm-trained group, follow-up dependent means  $t$  tests revealed that accuracy was higher for trained stimuli ( $M = 80.4\%, SEM = 2.01\%$ ) than for untrained stimuli ( $M = 71.7\%, SEM = 1.44\%$ ),  $t(15) = 4.89, p = .0002$ , Cohen's  $d = 1.20$ . In the control group, not surprisingly, accuracy was also found to be higher for trained stimuli ( $M = 66.7\%, SEM = 2.78\%$ ) than for untrained stimuli ( $M = 49.3\%, SEM = 0.82\%$ ),  $t(14) = 5.78, p < .0001$ , Cohen's  $d = 2.92$ . Follow-up independent means  $t$  tests revealed that accuracy was significantly higher in the algorithm-trained group than the control group for trained stimuli,  $t(29) = -4.03, p = .0004$ , Cohen's  $d = 1.70$ , and untrained stimuli,  $t(29) = -13.2, p < .0001$ , Cohen's  $d = 4.67$ . Unidirectional  $t$  tests were also carried out to determine whether accuracy was above chance for each



**Figure 4.** Experiment 1 behavioral results. (A) Average percent correct on the 2-AFC transfer test completed immediately after training on both trained (light gray) and untrained (dark gray) soundscapes for each group. (B) Average cross-modal accuracy in the triad task completed during the posttraining EEG recording session. Error bars denote standard deviation; chance performance for both tasks is 50% (indicated with dotted lines).



group and stimulus type separately. These tests confirmed that, as expected, accuracy in the control group was significantly above chance for trained stimuli,  $t(14) = 6.02, p < .0001$ , but not for untrained stimuli,  $t(14) = -0.08, p = .42$ . Importantly, accuracy in the algorithm-trained group was significantly above chance for both trained,  $t(15) = 15.1, p < .0001$ , and untrained (novel),  $t(15) = 15.0, p < .0001$ , stimuli.

Considering trials in which the auditory and visual stimulus matched as “target present” and those that mismatched as “target absent,” a mixed-model ANOVA with the within-participant factor Stimulus type (trained vs. untrained) and the between-participant factor Group (algorithm-trained vs. control) was used to test for differences in sensitivity ( $d'$ ). This ANOVA revealed main effects of Group,  $F(1, 29) = 50.51, p < .0001, \eta^2 = .55$ , and Stimulus type,  $F(1, 29) = 18.22, p = .0002, \eta^2 = .39$ , as well as a borderline significant interaction between Group and Stimulus type,  $F(1, 29) = 3.99, p = .0553, \eta^2 = .12$ . Follow-up independent means  $t$  tests revealed significant differences in  $d'$  between the algorithm-trained group (trained stimuli:  $d' = 1.58, SEM = 0.17$ ; untrained stimuli:  $d' = 1.22, SEM = 0.09$ ) and the control group (trained stimuli:  $d' = 0.95, SEM = 0.14$ ; untrained stimuli:  $d' = -0.03, SEM = 0.17$ ), for trained stimuli,  $t(29) = 2.90, p = .007$ , Cohen's  $d = 1.04$ , and for untrained stimuli,  $t(29) = 6.51, p < .0001$ , Cohen's  $d = 2.32$ . Dependent means  $t$  tests found that  $d'$  for trained stimuli was significantly higher than for untrained stimuli in both the algorithm-trained group,  $t(15) = 2.37, p = .03$ , Cohen's  $d = 0.76$ , and the control group,  $t(14) = 3.50, p = .0035$ , Cohen's  $d = 1.49$ .

RTs for the algorithm-trained group (trained stimuli:  $M = 555$  msec,  $SEM = 39$  msec; untrained stimuli:  $M = 622$  msec,  $SEM = 53$  msec) and the control group (trained stimuli:  $M = 517$  msec,  $SEM = 29$  msec; untrained stimuli:  $M = 519$  msec,  $SEM = 36$  msec) were also evaluated using a mixed-model ANOVA assessing the within-participant factor Stimulus type (trained vs. untrained) and the between-participant factor Group (algorithm-trained vs. control). This ANOVA revealed no significant main effects of Group,  $F(1, 29) = 1.73, p = .1989, \eta^2 = .29$ , or Stimulus type,  $F(1, 29) = 2.78, p = .1060, \eta^2 = .09$ , as well as no significant interaction,  $F(1, 29) = 2.52, p = .1236, \eta^2 = .08$ .

### Behavioral Results (Triad Task)

The triad task was performed during the pretraining and posttraining sessions while EEG was being recorded. This 2-AFC task required participants to determine if any two of the three stimuli matched (within- or across-modality). A ceiling effect was observed with regard to unimodal match accuracy, with near-perfect accuracy in the algorithm-trained group pretraining ( $M = 97.8\%$ ,  $SEM = 0.45\%$ ) and posttraining ( $M = 94.2\%$ ,  $SEM = 1.00\%$ ) and the same with the control group pretraining ( $M = 97.8\%$ ,  $SEM = 0.47\%$ ) and posttraining ( $M = 97.6\%$ ,  $SEM = 0.37$ ). Because cross-modal matches were irrelevant in the pretraining session

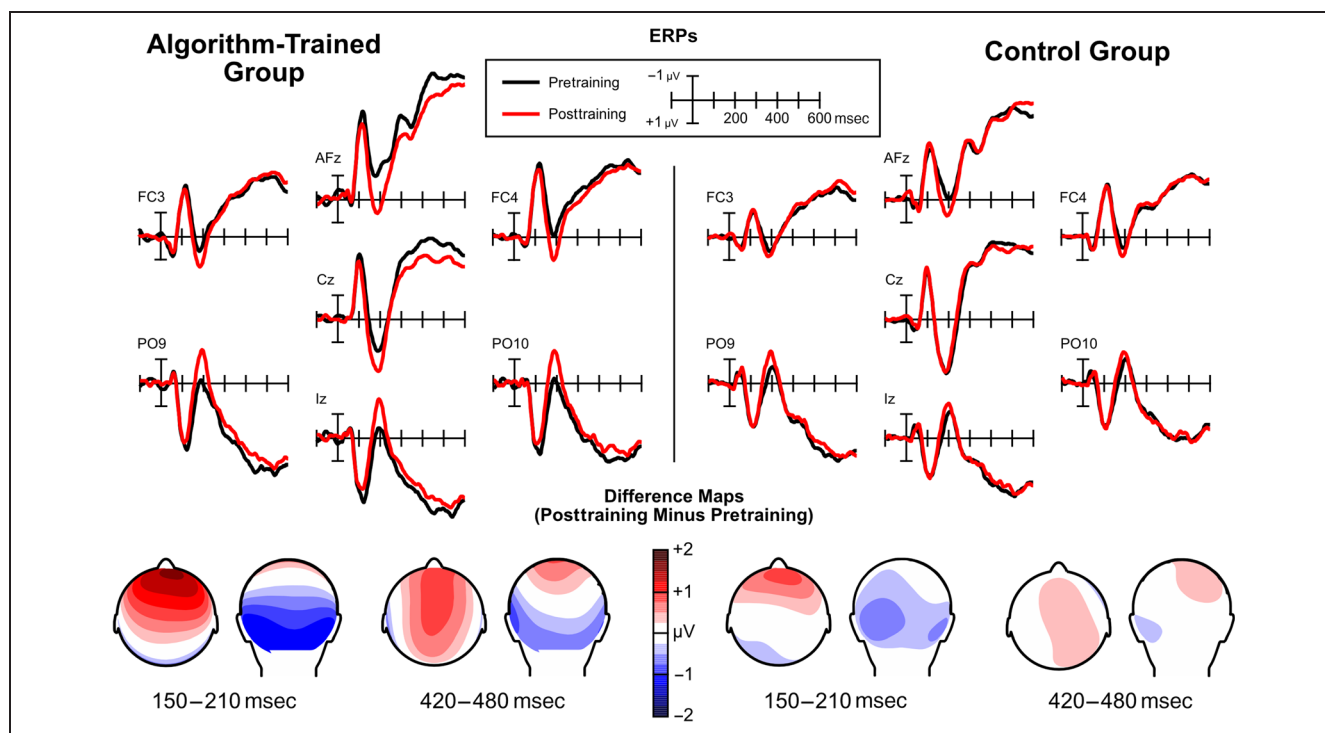
and both groups therefore performed at chance, we carried out two types of statistical tests on cross-modal accuracies only in the posttraining session (Figure 4B). First, unidirectional  $t$  tests for each group separately confirmed that cross-modal performance was above chance in the posttraining session for the algorithm-trained group ( $M = 74.6\%$ ,  $SEM = 1.76\%$ ),  $t(15) = 14.0, p < .0001$ , and for the control group ( $M = 65.9\%$ ,  $SEM = 2.43\%$ ),  $t(14) = 6.56, p < .0001$ . Second, between participants, independent  $t$  tests indicated enhanced cross-modal accuracy in the algorithm-trained group compared with the control group,  $t(29) = -2.94, p = .006$ , Cohen's  $d = 1.05$ . Note that RT analyses were not carried out on the triad task data, as the nature of this task (delayed response) does not allow for meaningful interpretation of these measures.

### ERP Results

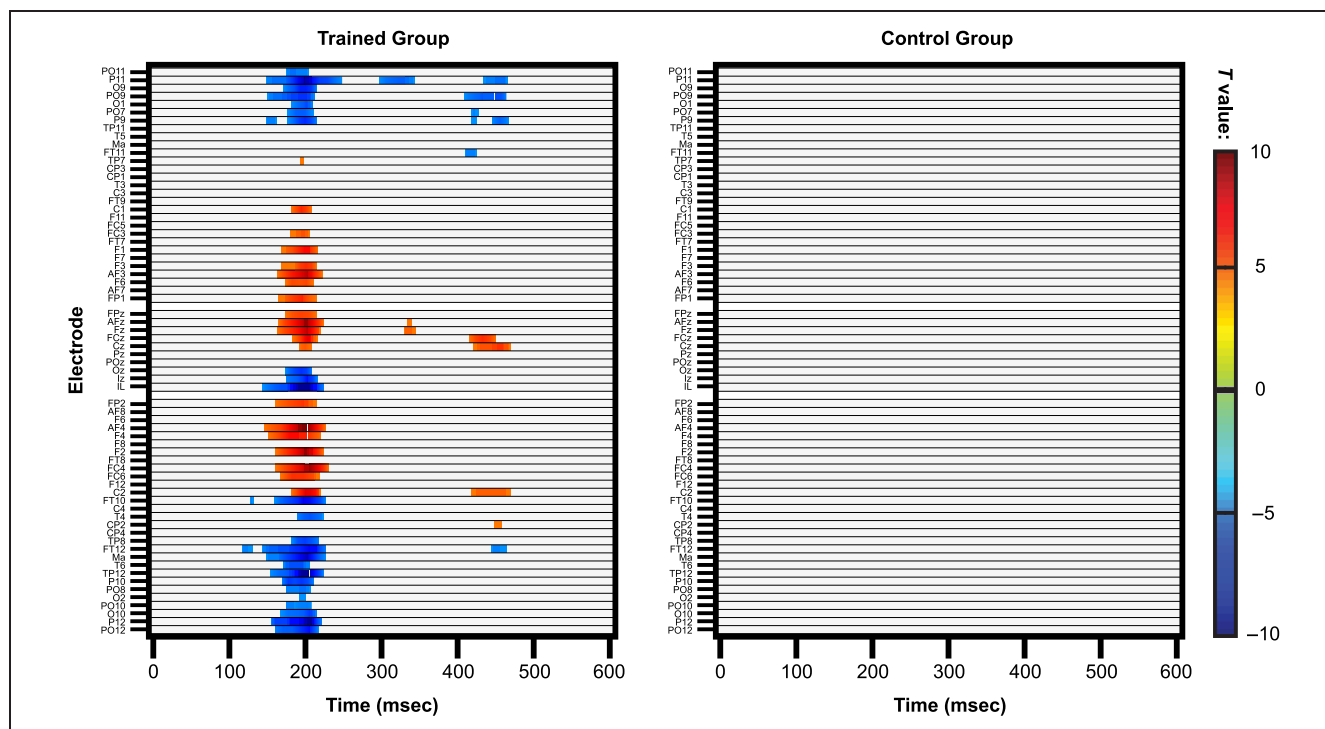
The main goal of our ERP analyses was to assess the time course of any neural changes that might have occurred because of sensory substitution training. To test this, we focused first on posttraining versus pretraining contrasts between ERPs elicited by the auditory soundscape stimuli (the first stimulus in each triad) in the algorithm-trained group and the control group separately (Figure 5). This contrast ensures that any observed ERP differences are most likely due to the intervening training, as the pretraining and posttraining ERPs were elicited by physically identical stimuli. We then compared the magnitude of any training-based ERP effect between the two groups.

The ERPs from pretraining and posttraining sessions were submitted to a conservative mass univariate analysis (MUA; carried out using ERP Toolbox; Groppe, Urbach, & Kutas, 2011), which was used in lieu of more conventional mean amplitude ANOVAs because we did not have a priori hypotheses for specific time windows and electrode locations for the ERP differences related to cross-modal perceptual learning. This analysis involved repeated-measures, two-tailed  $t$  tests at all time points between 0 and 600 msec (EEG was recorded at a 500-Hz sampling rate) at all 92 scalp electrodes (i.e., 27,600 total comparisons). The Benjamini and Yekutieli (2001) procedure for control of false discovery rate (FDR) was applied to assess the significance of each test using an FDR level of 5%. This ensures that the actual FDR (the mean proportion of significant test results that are actually false discoveries) will be less than or equal to the nominal FDR level of 5% regardless of the dependency structure of multiple comparisons.

In the algorithm-trained group, this analysis revealed two significant spatio-temporal clusters of posttraining minus pretraining ERP differences (FDR-corrected  $p < .05$ ): an early difference from 150 to 210 msec poststimulus and a later difference from 420 to 480 msec poststimulus (Figure 6). In contrast, the MUA revealed no significant posttraining versus pretraining differences in the control group. The first effect (150–210 msec) was characterized



**Figure 5.** Experiment 1 ERP results. Grand-averaged ERPs time-locked to soundscape onset before (black) and after (red) training in the algorithm-trained group (left;  $N = 16$ ) and the control group (right;  $N = 15$ ). Difference maps (bottom row) display the mean amplitude difference (posttraining minus pretraining) over the two significant time windows identified by the MUA (150–210 and 420–480 msec).



**Figure 6.** Experiment 1 ERP statistics. MUA of difference waves (posttraining minus pretraining ERPs) in the algorithm-trained group (left) and the control group (right). Significant differences appear as colored squares in the raster plots. Analysis was carried out at each time point from 0 to 600 msec poststimulus at all 92 scalp electrode sites. FDR analysis was used to correct for multiple comparisons (Groppe et al., 2011). For ease of viewing, this figure displays 68 of the 92 electrode sites included in the analysis, but note that all electrode sites showing a significant difference at any time point are included here.

by an enhanced negative-going wave around the time of the posterior N1 peak, along with a positive focus over anterior scalp regions. The second ERP difference (420–480 msec) was less spatially focused but showed the same general pattern of a posterior negative-going wave accompanied by an anterior–central positive wave. Scalp topography maps of these two ERP differences are provided in the bottom of Figure 5.

### Source Localization Results

A post hoc source localization analysis was performed to estimate possible sources of the ERP differences. Difference waves were calculated by subtracting the pretraining ERPs from the posttraining ERPs in each group. The low-resolution electromagnetic tomography method (Michel et al., 2004; Pascual-Marqui, Michel, & Lehmann, 1994) was applied to the resulting difference waves to localize possible cortical sources of the differential activity during the two time windows of interest (150–210 and 420–480 msec). Because we did not have a strong a priori hypothesis about the location of possible sources, we estimated activity levels of 42 ROIs across the whole brain, corresponding to the 42 Brodmann's areas (BAs). Permutation analyses were used to find ROIs with significantly above-zero activation (FDR level of 5%). Any ROIs that were found to be significant in both the algorithm-trained group and the control group are likely to reflect general changes due to training of sound–image pairing rather than changes specific to sensory substitution and thus were not analyzed further.

During the first time window of analysis (150–210 msec), four ROIs located in the inferior temporal lobe and inferior occipital lobe (BAs 37 and 42–44) showed significant activation in the algorithm-trained group, but not in the

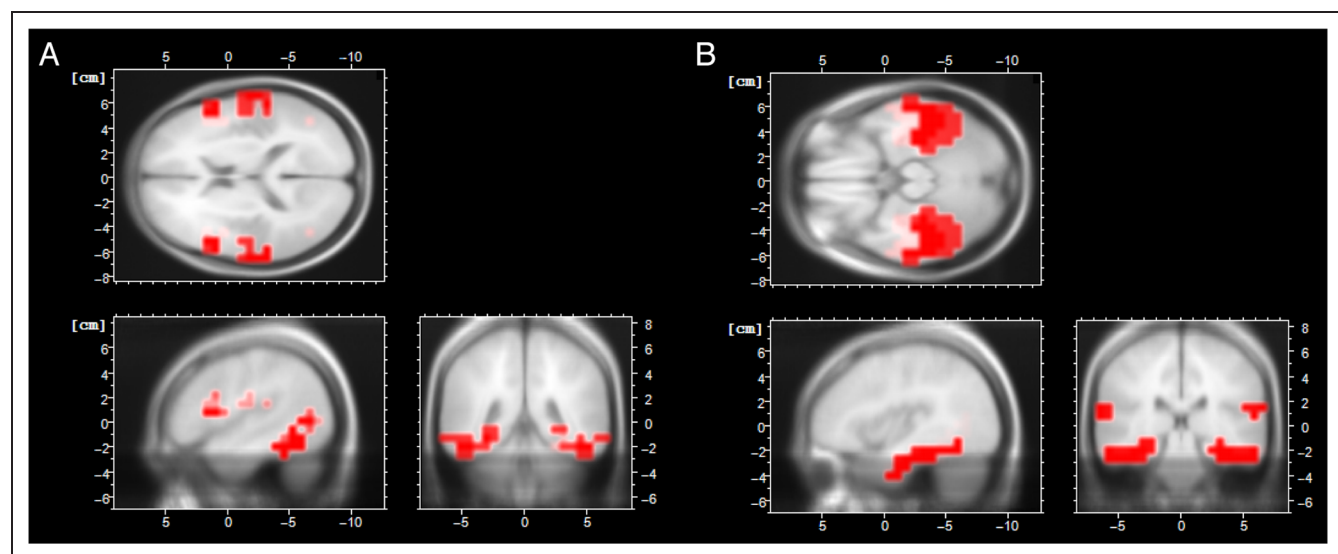
control group (see Figure 7). No significant ROIs specific to the control group were evident anywhere in the brain. During the second time window of analysis (420–480 msec), four ROIs, located in the temporal lobe and inferior occipital lobe (BAs 20, 36, 37, and 42), were found to have significant (FDR < 5%) activation in the algorithm-trained group, but not in the control group. Again, no ROIs were found to show significant activation in the control group alone.

### Discussion

The behavioral results suggested that training with the Meijer algorithm (for less than 2 hr in total) enabled auditory–visual sensory substitution. Not only did the algorithm-trained group perform better than the control group on the 80 stimuli they were trained on, but the algorithm-trained group was also able to successfully generalize their learning to 80 untrained stimuli. These results support previous findings that sensory substitution via the Meijer algorithm requires very little training to achieve a high level of accuracy (Kim & Zatorre, 2008).

The control group was able to perform above chance on the trained stimuli, with no apparent detriment to RTs, despite the fact that they had to memorize 80 arbitrary auditory–visual stimulus pairings. Note that this type of control group is considered to be ideal (in theory), but the task of memorizing random stimulus pairs was previously thought (in practice) to be too difficult for participants. Here, we demonstrate that having control participants rely solely on memory is a viable method for between-group comparisons in sensory substitution experiments.

Analysis of the auditory ERPs elicited by the soundscapes revealed two effects associated with training. The first was an early difference (posttraining minus



**Figure 7.** Source localization results showing significant ROIs at the 5% FDR level during (A) the first ERP effect (150–210 msec) and (B) the second ERP effect (420–480 msec) of Experiment 1. All ROIs shown were significant for the algorithm-trained group only. No control-group-specific significant ROIs were found.



pretraining) from 150 to 210 msec after soundscape onset. Although the neuroanatomical sources of this differential brain activity are difficult to estimate with ERPs alone, this negative-going wave over posterior scalp regions is consistent with generators in ventral visual cortical regions, perhaps reflecting visual cortex activation in response to the soundscape stimuli. On the basis of the results of our statistical analysis, it appeared that this early effect was absent in the control group. However, visual inspection of the ERPs and difference maps suggested that, although smaller, this early effect may also have been present in the control group (see Figure 5) but failed to reach the criteria for statistical significance using the conservative exploratory FDR-corrected analysis (Figure 6). A source localization analysis suggested that the changes in neural activity associated with sensory substitution training may have originated in the inferior temporal lobe and inferior occipital lobe, although these estimates should be interpreted with caution given the known limitations of such analyses.

The second training effect observed in the auditory ERPs was a difference at anterior and central electrode sites from 420 to 480 msec after soundscape onset. The timing of this effect would typically be considered “late” in terms of auditory sensory processing, but it is important to keep in mind that the soundscapes required 500 msec to sweep across the entire shape images. Again, our analysis revealed this effect to be present only in the algorithm-trained group, making it a possible candidate for a sensory substitution training effect. Our source localization analysis identified the medial/inferior temporal cortex and inferior occipital cortex as possible sources of this effect, which are broadly consistent with previous fMRI results (Kim & Zatorre, 2011; Amedi et al., 2007).

Despite these informative patterns in our ERP results, it is important to note the minor change in task between the posttraining and pretraining sessions. In the pretraining session, only unimodal matches were task relevant, whereas in the posttraining session, both unimodal and cross-modal matches were task relevant. Thus, it could be the case that the posttraining-versus-pretraining ERP effects index task-based attentional differences rather than processes specifically related to sensory substitution.

To test this possibility and replicate our initial findings, we carried out a second experiment with 36 new participants. In Experiment 2, we controlled for possible task-based differences by making the pretraining and posttraining tasks identical. That is, in both ERP sessions, participants were only asked to identify unimodal matches, whereas cross-modal matches remained task irrelevant. This change in the posttraining task also allowed us to assess whether the effects of sensory substitution training are evident even when the cross-modal information is irrelevant to the task. In addition, we sought to check whether our conservative FDR corrections and exploratory MUA may have missed any small (but genuine) differences in the control group’s posttraining versus

pretraining ERPs. Therefore, we increased our statistical power in Experiment 2 by restricting our analyses to the time windows identified as significant in Experiment 1. We also employed the more sensitive cluster permutation approach in our MUA instead of the exploratory FDR correction used in Experiment 1 (Groppe et al., 2011).

Thus, our main goals in Experiment 2 were to determine if we could replicate the findings of Experiment 1, assess whether the marginal (nonsignificant) ERP differences observed in the control group were genuine, and test whether the effects of sensory substitution training can be observed when the cross-modal information is irrelevant to the task.

## EXPERIMENT 2

### Methods

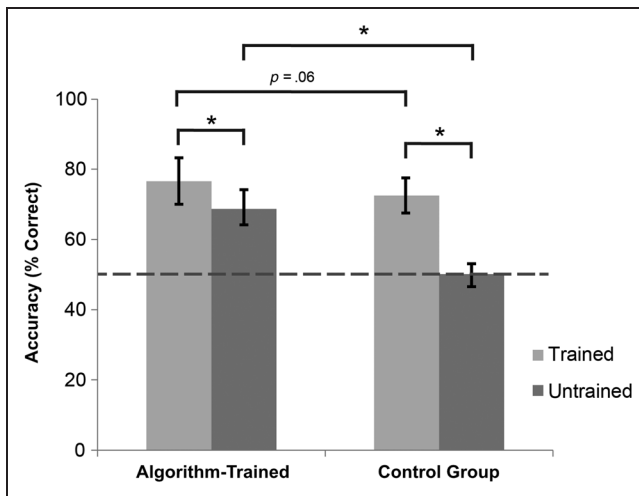
The design of Experiment 2 was nearly identical to that of Experiment 1. The participants included 36 college students (19 women), aged 19–24 (mean age = 20) years, with normal hearing and normal or corrected-to-normal vision. Again, the participants were split randomly into algorithm-trained and control groups. The data of four participants (two algorithm-trained and two control participants) were excluded from further analysis because greater than 30% of trials per individual were removed during artifact rejection. The remaining participants included 16 in the algorithm-trained group and 16 in the control group. On average, 17.4% of the trials were removed during artifact rejection in the algorithm-trained group, and 18.0% of the trials were removed from the control participants. The same stimuli were used; however, the randomized sound–image pairings for the control group in Experiment 1 were rerandomized for Experiment 2.

Two changes were made to the procedure. In both the pretraining and posttraining sessions, participants were tasked with identifying unimodal matches only. In short, the sound–image pairings from the intervening training were task irrelevant in both EEG sessions. In addition, half of the 80 auditory and visual stimuli presented in the posttraining session were untrained, meaning that participants had only been exposed to them in the pretraining session but not in the training session; that is, participants were trained on 40 stimuli but presented 80 stimuli in the EEG sessions (40 trained, 40 untrained). During the transfer test of Experiment 2, participants were tested on the 40 trained stimuli as well as on a separate set of 40 untrained stimuli that were not used in either the pretraining or posttraining sessions.

### Results

#### *Behavioral Results (Transfer Test)*

Just as in Experiment 1, participants in Experiment 2 completed a transfer test immediately after the three-part training (Figure 8). A mixed-model ANOVA assessing



**Figure 8.** Experiment 2 behavioral results. Average percent correct on the 2-AFC transfer test completed directly after training for both groups and on both trained and untrained stimuli. Error bars denote standard deviation; chance performance is 50% (indicated with the dotted line).

the within-participant factor Stimulus type (trained vs. untrained) and the between-participant factor Group (algorithm-trained vs. control) revealed main effects of Group,  $F(1, 30) = 43.2, p < .0001, \eta^2 = .20$ , and Stimulus type,  $F(1, 30) = 139.5, p < .0001, \eta^2 = .47$ , as well as an interaction between Group and Stimulus type,  $F(1, 30) = 20.3, p = .0001, \eta^2 = .09$ .

In the algorithm-trained group, follow-up dependent means  $t$  tests revealed that accuracy was higher for trained stimuli ( $M = 76.8\%$ ,  $SEM = 1.91\%$ ) than for untrained stimuli ( $M = 67.3\%$ ,  $SEM = 1.34\%$ ),  $t(15) = 5.78, p < .0001$ , Cohen's  $d = 1.44$ . In the control group, accuracy was also higher for trained stimuli ( $M = 72.0\%$ ,  $SEM = 1.47\%$ ) than for untrained stimuli ( $M = 50.8\%$ ,  $SEM = 0.99\%$ ),  $t(15) = 10.5, p < .0001$ , Cohen's  $d = 4.37$ . Follow-up independent means  $t$  tests revealed that accuracy was higher in the algorithm-trained group than the control group for untrained stimuli,  $t(30) = -9.88, p < .0001$ , Cohen's  $d = 3.54$ , and marginally higher for trained stimuli,  $t(30) = -1.98, p = .057$ , Cohen's  $d = 0.71$ . As in Experiment 1, we also carried out unidirectional  $t$  tests against chance. Accuracy in the algorithm-trained group was significantly above chance for both trained,  $t(15) = 14.0, p < .0001$ , and untrained stimuli,  $t(15) = 12.9, p < .0001$ . Accuracy in the control group was significantly above chance for trained stimuli,  $t(15) = 15.0, p < .0001$ , but not for untrained stimuli,  $t(15) = 0.83, p = .42$ .

Once again, a mixed-model ANOVA with the within-participant factor Stimulus type (trained vs. untrained) and the between-participant factor Group (algorithm-trained vs. control) was used to test for differences in sensitivity ( $d'$ ). This ANOVA revealed main effects of Group,  $F(1, 30) = 23.78, p < .0001, \eta^2 = .48$ , and Stimulus type,  $F(1, 30) = 21.94, p = .0001, \eta^2 = .42$ , as well as

a significant interaction between Group and Stimulus type,  $F(1, 30) = 9.47, p = .0044, \eta^2 = .24$ . Follow-up independent means  $t$  tests revealed no significant difference in  $d'$  between the algorithm-trained group ( $d' = 1.45$ ,  $SEM = 0.12$ ) and the control group ( $d' = 1.13$ ,  $SEM = 0.20$ ) for trained stimuli,  $t(30) = 1.36, p = .185$ , Cohen's  $d = 0.48$ , but there was a significant difference between the algorithm-trained group ( $d' = 1.20$ ,  $SEM = 0.17$ ) and the control group ( $d' = -0.08$ ,  $SEM = 0.14$ ) for untrained stimuli,  $t(30) = 5.94, p < .0001$ , Cohen's  $d = 2.10$ . Dependent means  $t$  tests found no significant difference in  $d'$  between trained and untrained stimuli in the algorithm-trained group,  $t(15) = 1.50, p = .156$ , Cohen's  $d = 0.43$ , but a significant difference in  $d'$  for trained versus untrained stimuli was observed in the control group,  $t(15) = 4.6, p = .0003$ , Cohen's  $d = 1.73$ .

Transfer test RTs for the algorithm-trained group (trained stimuli:  $M = 495$  msec,  $SEM = 33$  msec; untrained stimuli:  $M = 511$  msec,  $SEM = 30$  msec) and the control group (trained stimuli:  $M = 553$  msec,  $SEM = 28$  msec; untrained stimuli:  $M = 483$  msec,  $SEM = 37$  msec) were evaluated using a mixed-model ANOVA assessing the within-participant factor Stimulus type (trained vs. untrained) and the between-participant factor Group (algorithm-trained vs. control). This ANOVA revealed no significant main effects of Group,  $F(1, 30) = 0.5, p = .7009, \eta^2 = .01$ , or Stimulus type,  $F(1, 30) = 1.26, p = .2353, \eta^2 = .04$ , as well as no significant interaction,  $F(1, 30) = 3.38, p = .0761, \eta^2 = .07$ .

### Behavioral Results (Triad Task)

Unlike in Experiment 1, participants in Experiment 2 were only tasked with identifying unimodal matches during the triad task in both the pretraining and posttraining sessions. Unimodal accuracy on this task was at ceiling for the algorithm-trained group pretraining ( $M = 98.1\%$ ,  $SEM = 0.15\%$ ) and posttraining ( $M = 98.2\%$ ,  $SEM = 0.16\%$ ) as well as for the control group pretraining ( $M = 98.2\%$ ,  $SEM = 0.19\%$ ) and posttraining ( $M = 98.2\%$ ,  $SEM = 0.15\%$ ). RT and signal detection analyses were not carried out on the triad task data, as the nature of this task does not allow for meaningful interpretation of these measures.

### ERP Results

Auditory ERPs (Figure 9) elicited by the first stimulus delivered in each trial of the triad task posttraining versus pretraining were compared at two specific time windows: 130–230 and 400–500 msec. These time windows were selected a priori based on the timing of the two effects identified in Experiment 1 (150–210 and 420–480 msec; with  $\pm 20$  msec added to each window, as precise onsets/offsets of these effects are difficult to estimate). Because there were no observed differences between ERPs elicited by trained versus untrained stimuli, we collapsed across these stimuli for all subsequent analyses. The ERP

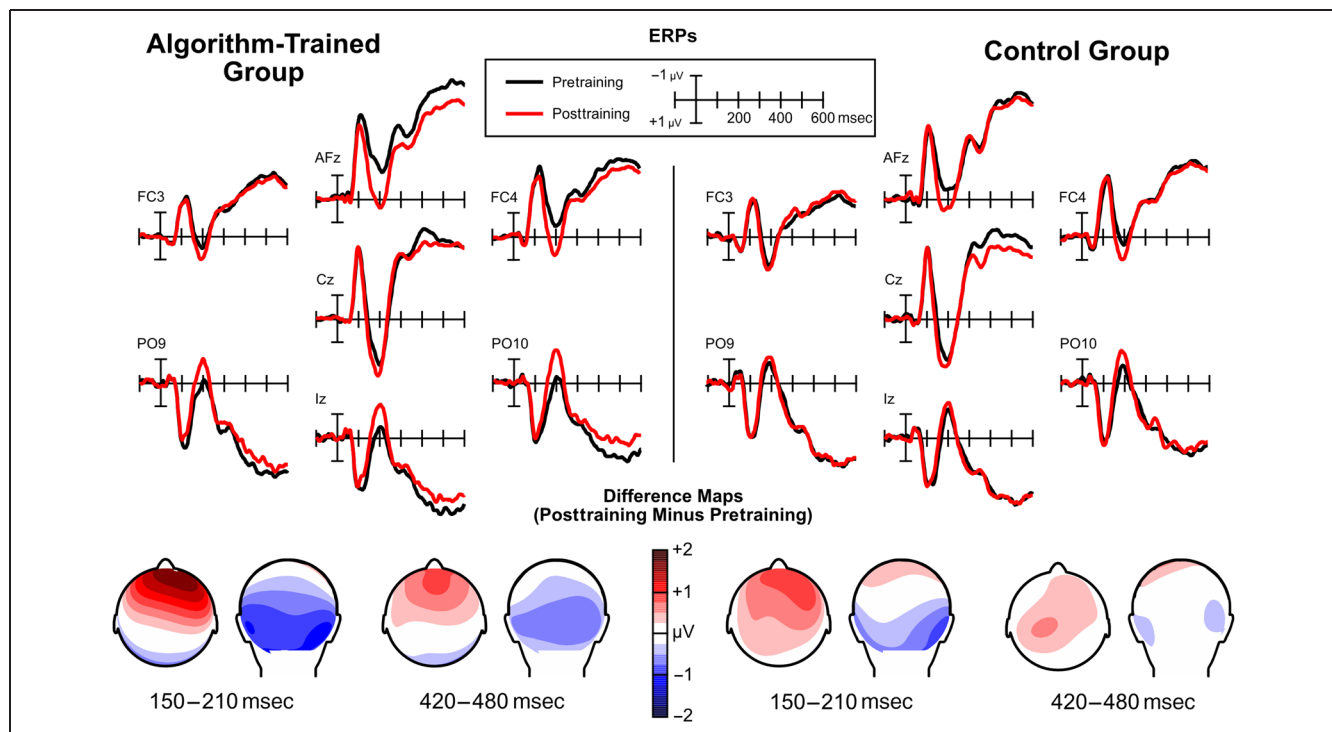
contrasts (posttraining vs. pretraining; within participants; identical stimuli, identical tasks) were analyzed using a cluster permutation approach (ERP Toolbox; Groppe et al., 2011). The results of this statistical analysis are provided in Figure 9.

Analysis within the first time window revealed a significant negative-going difference at posterior electrodes along with a positive difference at anterior/central electrodes, lasting from approximately 140 to 200 msec in both groups (Figures 9 and 10). Similar to Experiment 1, the low-resolution electromagnetic tomography method was employed to estimate possible cortical sources of this difference. Source localization analyses (FDR < 5%) revealed significant activation in the temporal lobe (BAs 20, 37, 42, and 43) and the dorsal posterior cingulate cortex (BA 31) that was specific to the algorithm-trained group (see Figure 11).

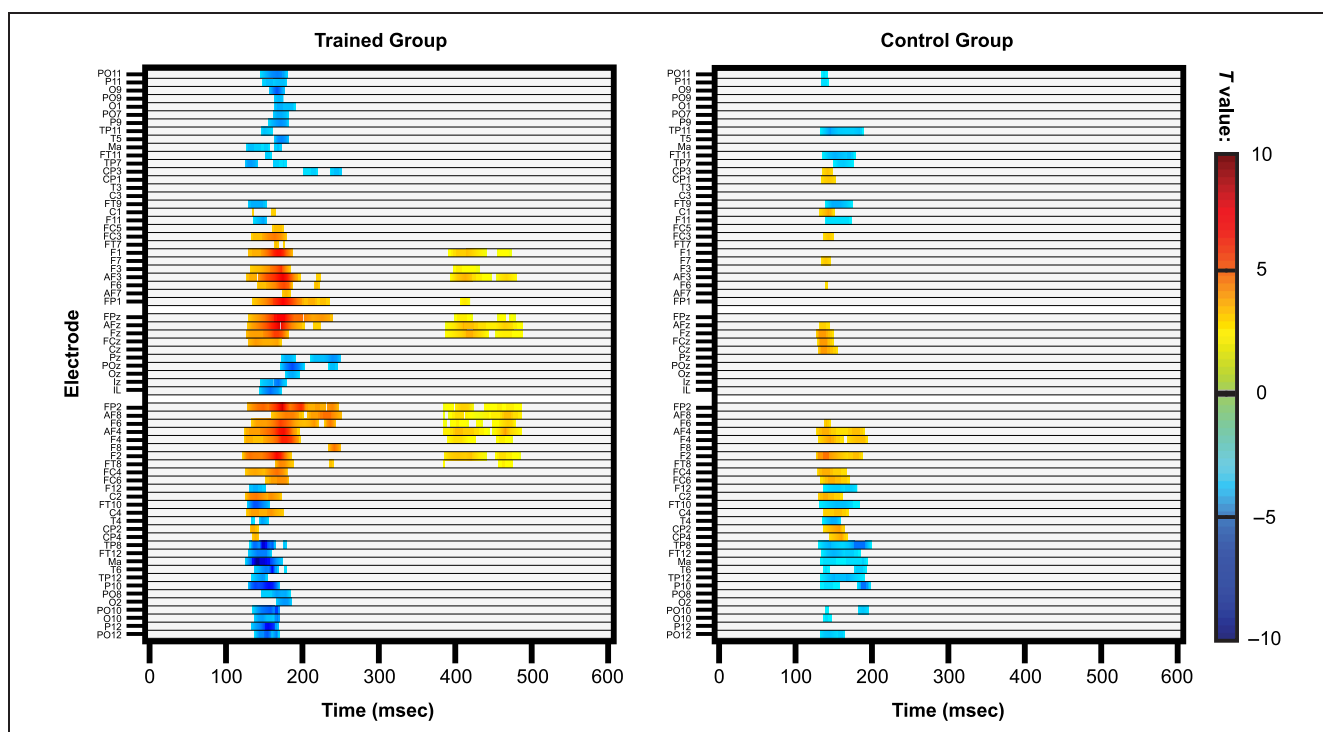
Analysis within the second time window revealed a significant positive difference at anterior–central electrode sites between 410 and 490 msec for the algorithm-trained group but not the control group. Source localization analyses suggested that this difference in activity may originate from the perirhinal cortex (BA 36) and the primary auditory cortex (BA 42). Significant activation was also found in the primary gustatory cortex (BA 43), although this is likely an artifact of its proximity to the auditory cortex and the low spatial precision of source localization of ERP signals. No significant sources specific to the control group were evident.

To test whether the early difference observed in both groups was larger in the algorithm-trained group compared with the control group, mean difference wave amplitudes across the relevant time window and electrodes (identified by the MUA) were compared using a one-way ANOVA. This early difference was evident in a wider distribution of electrodes in the algorithm-trained group compared with the control group (Figure 8). Therefore, to avoid biasing this test in favor of a significant difference between groups, a one-way ANOVA was used to compare the mean amplitude (150–210 msec) on pooled channels using only the electrodes that were found to be significantly different by the MUA in both groups during this time window. As this difference presented as a positivity at anterior electrodes and a negativity at posterior electrodes, the mean amplitudes taken from posterior electrodes were sign-inverted before pooling and subsequent analysis. This between-group comparison confirmed that the posttraining versus pretraining early ERP difference was significantly larger in the algorithm-trained group ( $M = 1.03 \mu\text{V}$ ,  $SEM = 0.18 \mu\text{V}$ ) compared with the control group ( $M = 0.48 \mu\text{V}$ ,  $SEM = 0.17 \mu\text{V}$ ),  $t(15) = -2.03$ ,  $p = .048$ ,  $\eta^2 = .14$ .

To test whether the effects observed in the algorithm-trained group were consistent across the two experiments (and thus potentially independent from the task), ANOVAs were used to compare the ERP differences in the algorithm-trained group between Experiments 1 and 2. Indeed, these analyses revealed no significant differences



**Figure 9.** Experiment 2 ERP results. Grand-averaged ERPs time-locked to soundscape onset before (black) and after (red) training in the algorithm-trained group (left;  $N = 16$ ) and the control group (right;  $N = 15$ ). Difference maps (bottom row) display the mean amplitude difference (posttraining minus pretraining) over the two significant time windows identified by the MUA (150–210 and 420–480 msec).

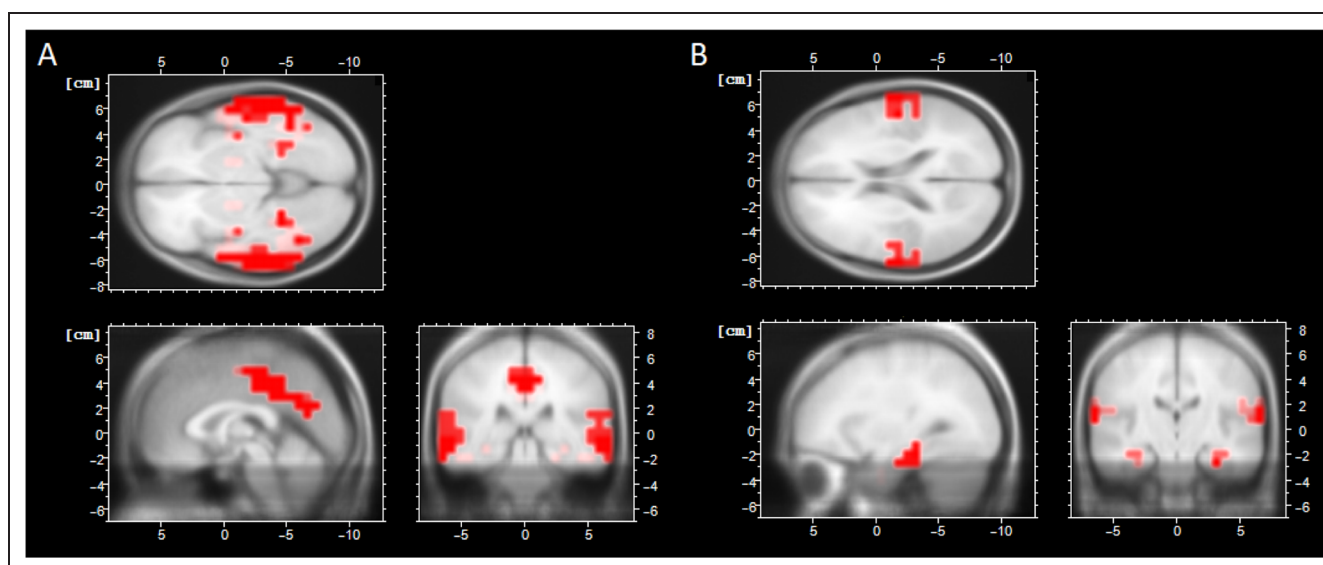


**Figure 10.** Experiment 2 ERP statistics. MUA of difference waves (posttraining minus pretraining ERPs) in the algorithm-trained group (left) and the control group (right). Significant differences appear as colored squares in the raster plots. Cluster permutation analyses (Groppe et al., 2011) were carried out at each time point within the two windows identified in Experiment 1. The first time window was from 150 to 210 msec poststimulus ( $\pm 20$  msec), and the second time window was from 420 to 480 msec poststimulus ( $\pm 20$  msec). Within these time windows, the analysis was carried out at all 92 scalp electrode sites. This figure displays 68 of the 92 electrode sites included in the analysis (for visualization purposes), but note that all electrode sites showing a significant difference at any time point are included here.

across experiments for both the early,  $F(1, 30) = 0.01, p = .91$ , and late,  $F(1, 30) = 2.04, p = .16$ , ERP time windows.

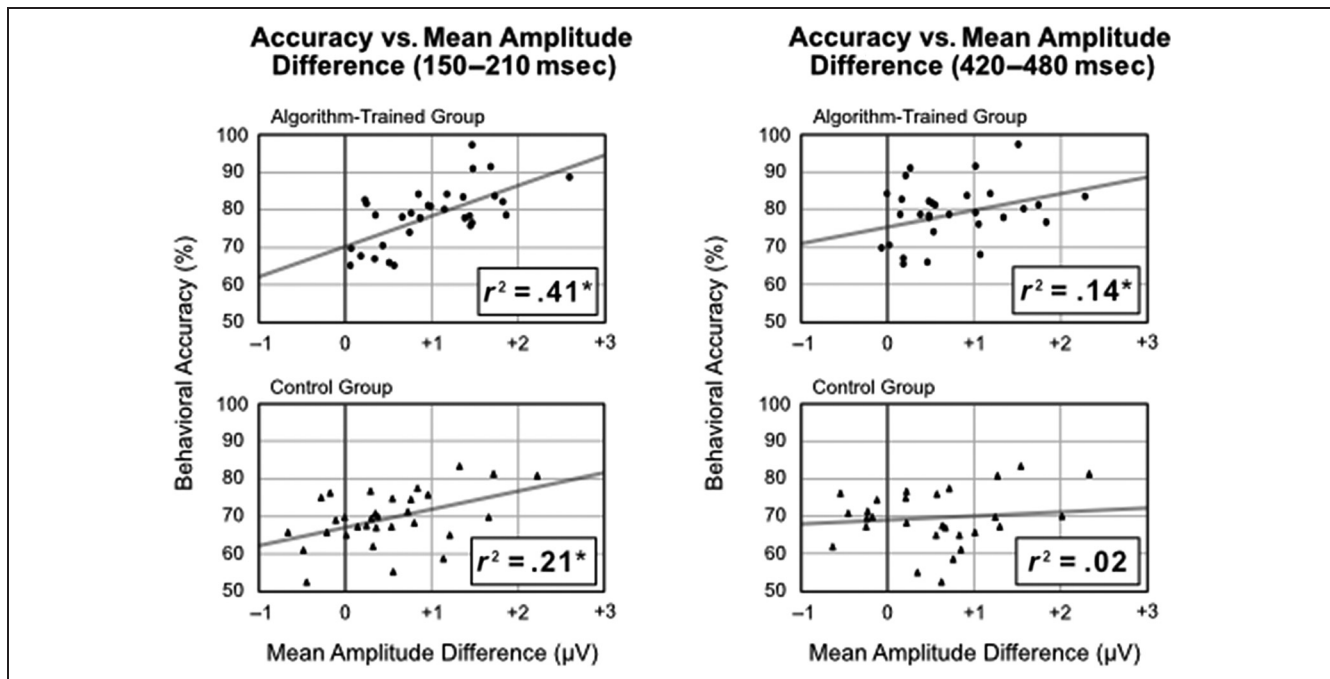
To further explore the functional significance of these ERP results, we conducted a series of correlational analyses between behavioral performance and the two ERP effects. First, we were interested in whether accuracy

on the behavioral tests administered immediately after the training (Session 2) correlated with the magnitude of the observed ERP differences (posttraining vs. pretraining), such that participants who exhibited better cross-modal learning would show larger ERP differences. For these analyses, we combined participants from



**Figure 11.** Source localization results showing significant ROIs at the 5% FDR level during (A) the first ERP effect (150–210 msec) and (B) the second ERP effect (420–480 msec) of Experiment 2. All ROIs shown were significant for the algorithm-trained group only. No control-group-specific significant ROIs were found.





**Figure 12.** Correlations between behavioral accuracy for trained stimuli and ERP differences (posttraining minus pretraining) for the algorithm-trained group (top row) and the control group (bottom row), during the early ERP time window (left column) and the late ERP time window (right column). Data were collapsed across Experiments 1 and 2. Percent accuracy, which was tested directly after training in Session 2, is plotted against posttraining (Session 3) minus pretraining (Session 1) ERP differences. Asterisks indicate significant correlations. See text for details on how the ERP amplitudes were quantified for these correlations.

Experiments 1 and 2 to increase statistical power. To quantify the ERP differences, we once again utilized mean amplitudes over the early and late time windows at pooled channels using electrodes identified by the MUA (sign-inverted for posterior negative-going effects). We first compared ERP differences against accuracy for trained stimuli (Figure 12). Higher accuracy on the behavioral test positively correlated with differential ERP magnitudes for the algorithm-trained group during both the early, Pearson's  $r(32) = .64, p < .01$ , and late, Pearson's  $r(32) = .38, p = .03$ , time windows. A positive correlation was also present in the control group during the early time window, Pearson's  $r(31) = .46, p < .01$ , but not during the later time window, Pearson's  $r(31) = .046, p = .47$ . Correlation analyses were also carried out using behavioral accuracy for untrained stimuli. In the algorithm-trained group, trending positive correlations were observed during the early window, Pearson's  $r(32) = .32, p = .07$ , and the late window, Pearson's  $r(32) = .13, p = .05$ . In the control group, no significant correlations were observed during either time window (early: Pearson's  $r(31) = .01, p = .60$ ; late:  $r(31) = -.09, p = .62$ ).

We were also interested in whether behavioral performance during the EEG sessions would correlate with the observed ERP effects. Such analyses were only possible for Experiment 1 in which the cross-modal matches were task relevant (i.e., In Experiment 2, behavioral performance was at ceiling for both groups because only unimodal matches were required by the task). Here, we

observed a positive correlation between cross-modal accuracy and differential ERP magnitude in the algorithm-trained group during the early time window only, Pearson's  $r(16) = .57, p = .03$ , with no significant correlation during the late time window, Pearson's  $r(16) = -.15, p = .59$ . In the control group, no significant correlations were evident in the early, Pearson's  $r(15) = .39, p = .16$ , or late, Pearson's  $r(15) = -.06, p = .83$ , time windows.

## GENERAL DISCUSSION

In both experiments, sensory substitution training with the Meijer algorithm for a mere 2 hours was sufficient for sighted individuals to learn to identify complex visual shapes from their corresponding auditory soundscapes. Furthermore, these cross-modal perceptual learning effects were generalizable, such that participants in the algorithm-trained group were able to identify novel visual shapes via novel soundscapes at well above-chance levels (72% in Experiment 1, 67% in Experiment 2; in both cases, chance = 50%). We estimated the time course of auditory–visual sensory substitution by isolating ERP differences (posttraining vs. pretraining) in the group of participants trained with the Meijer algorithm, compared with a tightly matched control group. An early training-based ERP modulation (150–210 msec), with a focus over the posterior scalp, was significantly enhanced in the algorithm-trained group. A subsequent ERP difference (420–480 msec) was present only in the algorithm-trained group. Both of



these ERP effects occurred relatively early in time, considering that the soundscape took 500 msec to sweep across the image from left to right. The timing and scalp distribution of these ERP effects are suggestive of fairly rapid visual cortex activation in response to the auditory soundscapes.

There are a few possible interpretations of the earliest ERP effect observed here (150–210 msec). First, an increase in task difficulty (and greater cognitive effort) for the algorithm-trained group can be ruled out, as the post-training task in Experiment 1 was almost certainly more difficult for the control group (who were tasked with memorizing 80 arbitrary stimulus pairings) compared with the algorithm-trained group (who could rely on a few simple rules from the algorithm to complete the task). Importantly, this same early ERP modulation was replicated in Experiment 2, despite the cross-modal pairings remaining task-irrelevant in the posttraining session. Given the larger magnitude of this ERP modulation in the algorithm-trained group, one could argue that this neural difference reflects an increase in auditory attention in both groups (due to the intervening training) but that the algorithm-trained group had a relatively greater attentional load because they learned the rules of the algorithm in addition to the individual stimulus pairings. In Experiment 2, however, we cannot rule out the possibility that the algorithm-trained group differentially attended to the soundscapes and continued to extract shape information from the auditory stimuli even when cross-modal pairings were task irrelevant.

A related interpretation is that the early ERP difference observed here is analogous to the training-related auditory P2 effect observed in many previous studies investigating auditory perceptual learning (Tremblay, Ross, Inoue, McClannahan, & Collet, 2014; Tong, Melara, & Rao, 2009; Trainor, Shahin, & Roberts, 2003; Tremblay, Kraus, McGee, Ponton, & Otis, 2001). In these studies, participants were initially exposed to auditory stimuli that were difficult to discriminate and were then trained to discriminate the stimuli. The resulting posttraining versus pre-training ERP difference typically appears as an early positivity at anterior/central scalp electrode sites, overlapping the evoked auditory P2 peak (Tremblay et al., 2001, 2014; Tong et al., 2009; Trainor et al., 2003). This auditory P2 effect has been hypothesized to reflect neuroplastic changes in the auditory cortex as a result of auditory perceptual learning, so it would be reasonable to observe such an effect here (Trainor et al., 2003). The scalp distribution of the early ERP difference observed in the current experiments, however, does not appear to be consistent with typical auditory attention-based modulations or unimodal auditory perceptual learning effects.

Post hoc analyses were conducted to test whether the scalp distribution of the early effect differed from the distribution of the auditory-evoked P2 component in the algorithm-trained group. Amplitudes were normalized (McCarthy & Wood, 1985) and compared via ANOVA with

the factors ERP component (early training-related difference vs. auditory-evoked P2) and Scalp channel location (1–96). This analysis showed a significant interaction between ERP component and Scalp channel location, suggesting that the scalp distribution of the training-related effect was indeed distinct from that of the auditory-evoked P2,  $F(95, 1425) = 17.69$ ,  $p < .0001$ ,  $\eta^2 = .22$ . So although this early ERP difference may, in part, reflect a modulation in the auditory P2, the scalp distribution seems to suggest that there are additional activity changes in other brain regions as a result of the training.

As the scalp distribution of the training-related effect suggests possible occipital sources, one interpretation is that the early ERP difference may reflect visual cortex activity as a result of the SSD training. The presence of a similar (but much smaller) ERP difference in the control group does not necessarily discount this interpretation, as we intentionally matched the audiovisual training in the control group as tightly as possible. The difference in magnitude of this effect between the algorithm-trained group and the control group might index enhanced cross-sensory processing afforded by the utilization of the rules of the Meijer algorithm. In other words, both groups learned to associate sounds to shapes (hence the presence of this difference in both groups), but the algorithm-trained group learned to extract shape information based on the image-to-sound conversion rules, resulting in the substantial increase in the magnitude of the effect. The results of the source localization analyses are consistent with this interpretation, with sources in medial/inferior temporal and occipital cortex found to be uniquely associated with sensory substitution (in the algorithm-trained group) rather than more generic auditory-to-visual association learning (in both groups). Future studies could be carried out using both ERPs and fMRI to determine if a connection is present between the ERP effect observed here and the enhanced functional connectivity between auditory cortex and ventral visual cortical areas, including the LOC observed by Kim and Zatorre (2011). In such a study, it would be useful to train participants over longer periods in multiple training sessions with intervening ERP/fMRI sessions to more precisely link the spatio-temporal neuroplastic changes to the cross-sensory learning process as it develops (e.g., see Ward & Meijer, 2010).

The ERP difference observed from 420 to 480 msec in the algorithm-trained group was absent in the control group, despite the fact that they were presented with identical stimuli, received the same amount of training, and performed the same tasks in each session. In addition, this later effect replicated with a new set of participants using the same basic paradigm in Experiment 2. On the basis of these observations, this longer latency ERP effect may also be related to auditory–visual sensory substitution. Previous studies of sensory substitution have used the fMRI technique and found that training can

change the magnitude and functional connectivity of BOLD responses in a modality-independent shape processing region, the LOC, as well as several other non-primary sensory areas such as the left precentral sulcus and the right occipito-parietal sulcus (Kim & Zatorre, 2011; Striem-Amit et al., 2011; Amedi et al., 2005, 2007). Because of the sluggish nature of fMRI, however, it was unclear whether these neural changes were due to a rapid (direct) transfer of sensory information or to a slower (indirect) cognitive route. The ERP effects observed here (especially the early difference with a posterior distribution from 150 to 210 msec) are consistent with the rapid transfer view of sensory substitution, in which auditory information is directly routed to shape processing areas after training with SSDs. Although these findings generally support the rapid transfer view of sensory substitution, they do not exclude the possibility that some neural regions may be involved in a slower, more indirect manner, and it is possible that the later ERP effect observed here indexes a cognitive component of the cross-modal learning process. These interpretations are limited by the poor spatial resolution of ERPs. Although source localization in ERP studies should be interpreted with caution (a particular scalp distribution may be generated by any number of sources in any number of locations), the observation of consistent activation in the medial and inferior temporal lobe, as well as the inferior occipital lobe, leaves open the possibility of a direct translation of the auditory signal in the auditory cortex to a more broad multisensory signal across inferior temporal and occipital cortices.

Finally, we observed positive correlations between behavioral performance on the cross-modal matching tasks and the magnitude of the ERP effects. These correlations were stronger in the algorithm-trained group compared with the control group and were particularly robust for the early ERP difference (150–210 msec). These brain-behavior relationships suggest that individuals who learned the best tended to show the largest neural changes, and vice versa. Smaller (and, in some cases, absent) correlations during the late ERP time window further suggest a closer link between the early ERP effect and audiovisual perceptual learning.

Overall, the electrophysiological effects of sensory substitution identified in the current study complement results from the fMRI literature and support the view that sensory substitution likely involves a fairly rapid transfer of sensory information. The identification of these training-based ERP differences provides several opportunities for future investigation, as more experiments must be carried out to determine what factors may influence the timing, magnitude, and presence of the observed neural changes. There also remains much work to be done to bridge the gap between the temporally sensitive ERP information gained from this study and the spatially sensitive information gained from fMRI studies.

## Conclusion

Patterns of electrophysiological brain activity that appear to be a direct result of sensory substitution training were identified in two separate experiments. These effects, an early ERP difference from 150 to 210 msec and a subsequent difference from 420 to 480 msec, are consistent with the view that sensory substitution may involve rapid transfer of information from one sensory modality to another. The current study also demonstrates that using arbitrary associations with a large stimulus set (80 sound–image pairs) is a viable method for establishing a matched control group in this area of research. Overall, our findings provide important temporal information that complements results of previous studies using spatially precise but temporally sluggish techniques and brings us one step closer to understanding the proposed neuroplastic networks that support auditory–visual sensory substitution.

## Acknowledgments

This work was supported by the Murdock Charitable Trust Research Program for Life Sciences as well as the Reed College Science Research Fellowship.

Reprint requests should be sent to Christian Gaulty, Department of Psychology, Reed College, 3203 S. E. Woodstock Blvd., Portland, OR 97202, or via e-mail: cgraulty@gmail.com.

## REFERENCES

- Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, 4, 324–330.
- Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron*, 48, 859–872.
- Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, 10, 687–689.
- Arno, P., Capelle, C., Wanet-Defalque, M. C., Catalan-Ahumada, M., & Veraart, C. (1999). Auditory coding of visual patterns for the blind. *Perception*, 28, 1013–1029.
- Arno, P., De Volder, A. G., Vanlierde, A., Wanet-Defalque, M.-C., Streel, E., Robert, A., et al. (2001). Occipital activation by pattern recognition in the early blind using auditory substitution for vision. *Neuroimage*, 13, 632–645.
- Bach-y-Rita, P., Collins, C. C., Sanders, F. A., White, B., & Scadden, L. (1969). Visual substitution by tactile image projection. *Nature*, 221, 963–964.
- Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., & Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: A technical note. *Journal of Rehabilitation Research and Development*, 35, 427–430.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29, 1165–1188.
- Bermejo, F., Di Paolo, E. A., Hug, M. X., & Arias, C. (2015). Sensorimotor strategies for recognizing geometrical shapes: A comparative study with different sensory substitution devices. *Frontiers in Psychology*, 6, 679.

- Bubic, A., Striem-Amit, E., & Amedi, A. (2010). Large-scale brain plasticity following blindness and the use of sensory substitution devices. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory object perception in the primate brain* (pp. 351–380). New York: Springer.
- Bueti, D., & Macaluso, E. (2010). Auditory temporal expectations modulate activity in visual cortex. *Neuroimage*, *51*, 1168–1183.
- Campus, C., Brayda, L., De Carli, F., Chellali, R., Fama, F., Bruzzo, C., et al. (2012). Tactile exploration of virtual objects for blind and sighted people: The role of beta 1 EEG band in sensory substitution and supramodal mental mapping. *Journal of Neurophysiology*, *107*, 2713–2729.
- Collignon, O., Lassonde, M., Lepore, F., Bastien, D., & Veraart, C. (2006). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cerebral Cortex*, *17*, 457–465.
- Cronly-Dillon, J., Persaud, K., & Gregory, R. P. F. (1999). The perception of visual images encoded in musical form: A study in cross-modality information transfer. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, *266*, 2427–2433.
- Driver, J., & Noesselt, T. (2007). Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron*, *57*, 11–23.
- Feng, W., Störmer, V. S., Martinez, A., McDonald, J. J., & Hillyard, S. A. (2014). Sounds activate visual cortex and improve visual discrimination. *Journal of Neuroscience*, *34*, 9817–9824.
- Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). fMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. *Neuroimage*, *54*, 1231–1243.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409–1422.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, *21*, 191–202.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725.
- Hertz, U., & Amedi, A. (2015). Flexibility and stability in sensory processing revealed using visual-to-auditory sensory substitution. *Cerebral Cortex*, *25*, 2049–2064.
- Kim, J.-K., & Zatorre, R. J. (2008). Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Research*, *1242*, 263–275.
- Kim, J.-K., & Zatorre, R. J. (2011). Tactile-auditory shape learning engages the lateral occipital complex. *Journal of Neuroscience*, *31*, 7848–7856.
- Levy-Tzedek, S., Hanassy, S., Abboud, S., Maidenbaum, S., & Amedi, S. (2012). Fast, accurate reaching movements with a visual-to-auditory sensory substitution device. *Restorative Neurology and Neuroscience*, *30*, 313–323.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
- Meijer, P. B. L. (1992). An experimental system for auditory image representation. *Transactions on Biomedical Engineering*, *39*, 111–121.
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*, 2195–2222.
- Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia*, *83*, 161–169.
- Ortiz, T., Poch, J., Santos, J. M., Requena, C., Martínez, A. M., Ortiz-Terán, L., et al. (2011). Recruitment of occipital cortex during sensory substitution training linked to subjective experience of seeing in people with blindness. *PLoS One*, *6*, e23264.
- Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*, 49–65.
- Peltier, S., Stilla, R., Mariola, E., LaConte, S., Hu, X., & Sathian, K. (2007). Activity and effective connectivity of parietal and occipital cortical regions during haptic shape perception. *Neuropsychologia*, *45*, 476–483.
- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbi, M. I., Wu, W.-H. C., Cohen, L., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences, U.S.A.*, *101*, 5658–5663.
- Sampaio, E., Maris, S., & Bach-y-Rita, P. (2001). Brain plasticity: “Visual” acuity of blind persons via the tongue. *Brain Research*, *908*, 204–207.
- Striem-Amit, E., Dakwar, O., Reich, L., & Amedi, A. (2011). The large-scale organization of “visual” streams emerges without visual experience. *Cerebral Cortex*, *22*, 1698–1709.
- Striem-Amit, E., Guendelman, M., & Amedi, A. (2012). “Visual” acuity of the congenitally blind using visual-to-auditory sensory substitution. *PLoS One*, *7*, e33136.
- Tong, Y., Melara, R. D., & Rao, A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Research*, *1297*, 80–88.
- Trainor, L., Shahin, A., & Roberts, L. (2003). Effects of musical training on the auditory cortex in children. *Annals of the New York Academy of Sciences*, *999*, 506–513.
- Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, *8*, 28.
- Tremblay, K. L., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1–P2 complex after speech-sound training. *Ear & Hearing*, *22*, 79–90.
- Ward, J., & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, *19*, 492–500.