

# Neural time course of visually enhanced echo suppression

Christopher W. Bishop,<sup>1</sup> Sam London,<sup>1</sup> and Lee M. Miller<sup>1,2</sup>

<sup>1</sup>Center for Mind and Brain, University of California, Davis, California; and <sup>2</sup>Department of Neurobiology, Physiology, and Behavior, University of California, Davis, California

Submitted 27 February 2012; accepted in final form 8 July 2012

**Bishop CW, London S, Miller LM.** Neural time course of visually enhanced echo suppression. *J Neurophysiol* 108: 1869–1883, 2012. First published July 11, 2012; doi:10.1152/jn.00175.2012.—Auditory spatial perception plays a critical role in day-to-day communication. For instance, listeners utilize acoustic spatial information to segregate individual talkers into distinct auditory “streams” to improve speech intelligibility. However, spatial localization is an exceedingly difficult task in everyday listening environments with numerous distracting echoes from nearby surfaces, such as walls. Listeners’ brains overcome this unique challenge by relying on acoustic timing and, quite surprisingly, visual spatial information to suppress short-latency (1–10 ms) echoes through a process known as “the precedence effect” or “echo suppression.” In the present study, we employed electroencephalography (EEG) to investigate the neural time course of echo suppression both with and without the aid of coincident visual stimulation in human listeners. We find that echo suppression is a multistage process initialized during the auditory N1 (70–100 ms) and followed by space-specific suppression mechanisms from 150 to 250 ms. Additionally, we find a robust correlate of listeners’ spatial perception (i.e., suppressing or not suppressing the echo) over central electrode sites from 300 to 500 ms. Contrary to our hypothesis, vision’s powerful contribution to echo suppression occurs late in processing (250–400 ms), suggesting that vision contributes primarily during late sensory or decision making processes. Together, our findings support growing evidence that echo suppression is a slow, progressive mechanism modifiable by visual influences during late sensory and decision making stages. Furthermore, our findings suggest that audiovisual interactions are not limited to early, sensory-level modulations but extend well into late stages of cortical processing.

auditory spatial perception; audiovisual integration; precedence effect; echo suppression

AUDIOVISUAL INTEGRATION plays a pivotal role in day-to-day communication. For instance, listeners rely on visual cues like a speaker’s mouth movements to better understand a talker in noisy environments (Sumbly and Pollack 1954). In some cases, vision’s influence over audition can be powerful enough to change where and what we hear. The ventriloquist’s illusion is a classic example of vision’s dominance over where we hear sounds in space (Howard and Templeton 1966). The illusion is evoked by presenting listeners with spatially disparate but temporally coincident auditory and visual stimuli. When listeners indicate the perceived location of the sound, their responses are strongly biased toward, and in some cases completely captured by, the visual stimulus (De Gelder and Bertelson 2003; Howard and Templeton 1966; Recanzone 2009; Thomas 1941; Witkin et al. 1952). Recently, we demonstrated that vision can further contribute to auditory spatial perception by helping listeners identify and perceptually suppress distracting echoes (Bishop et al. 2011). To do this, we employed

an auditory spatial illusion called the precedence effect or “echo suppression,” a phenomenon in which listeners spatially suppress echoes arriving shortly after the “precedent” sound wave (1–10 ms) but continue to hear long-latency (>10 ms) echoes (Litovsky et al. 1999; Wallach et al. 1949). Interestingly, much of the echo’s nonspatial information (e.g., intensity, timbre, and even temporal gaps) is largely preserved, suggesting that spatial information is selectively altered during echo suppression (Li et al. 2005; Litovsky et al. 1999). When listeners experience “echo suppression,” they perceive a single sound image at or near the location of the temporally leading sound that possesses properties of both the precedent sound wave and its echo. Although this is putatively believed to be an early, auditory-only phenomenon, the report by Bishop et al. (2011) provides psychophysical evidence that a coincident visual stimulus at the location of the precedent sound wave enhances echo suppression by as much as 20%. Here we used electroencephalography (EEG) to investigate the time course of this visual enhancement in human listeners.

The precedence effect and audiovisual integration are strongly linked to cortical processing. Despite considerable evidence in animal models implicating subcortical (Litovsky et al. 1999) and cortical (Cranford et al. 1971; Cranford and Oberholtzer 1976; Whitfield et al. 1978) neural substrates in the precedence effect, human EEG studies have been largely unsupportive of subcortical involvement (Damaschke et al. 2005; Liebenthal and Pratt 1999). Instead, recent studies have demonstrated that the precedence effect first manifests in primary auditory cortex (Liebenthal and Pratt 1997, 1999) with neural correlates that persist for hundreds of milliseconds after stimulus onset (Backer et al. 2010; Damaschke et al. 2005; Dimitrijevic and Stapells 2006; Sanders et al. 2008, 2011). Together, these findings suggest that the precedence effect is a slow, progressive process with ample opportunity for cross-modal modification. In contrast to the strong links between the precedence effect and cortical processing in humans, imaging and electrophysiological studies suggest that audiovisual interactions are ubiquitous throughout subcortical and cortical regions, including early sensory- and higher-level association areas (De Gelder and Bertelson 2003; Driver and Noesselt 2008; Ghazanfar and Chandrasekaran 2007; Macaluso 2006; Schroeder et al. 2003). This has led to a radical shift from a belief in largely insulated unisensory processing through early cortical stages toward a view that the human brain is fundamentally multisensory (Ghazanfar and Schroeder 2006). In light of strong evidence demonstrating neural correlates of both the precedence effect (i.e., echo suppression) and audiovisual interactions throughout cortex, we hypothesized that the previously described visual enhancement of echo suppression (Bishop et al. 2011) would manifest as changes in early auditory representations, as early as the auditory N1. In the present study, we tested this hypothesis by coupling a

Address for reprint requests and other correspondence: C. W. Bishop, Center for Mind and Brain, Univ. of California, 267 Cousteau Pl., Davis, CA 95618 (e-mail: cwbishop@ucdavis.edu).

variant of our previously described paradigm with the precise temporal information provided by EEG.

## METHODS

**Subjects.** In accordance with policies and procedures approved by the University of California, Davis Institutional Review Board, 39 subjects were recruited to participate in the present experiment [23 women, 16 men; mean  $22 \pm 3.32$  (SD) yr of age]. Subjects provided written consent prior to participating and were monetarily compensated for their time. Twelve subjects were excused prior to the EEG experiment as a result of poor performance or echo thresholds that fell below a preestablished cutoff of 2 ms during the behavioral prescreening session (see below for details). Of the remaining 27 individuals, 4 were excused during the EEG session because the experimenter failed to establish a stable psychophysical parameter (“echo threshold,” see below) during testing. An additional two subjects were excluded from the final analysis because of inadequate trial numbers in one or more conditions ( $<50$  prior to artifact rejection, which was below our preestablished cutoff of at least 100 trials for each condition). Thus the analyses reported here are based on the remaining 21 individuals [12 women, 9 men; mean  $21 \pm 1.89$  (SD) yr of age]. These subjects were divided into two groups differing in the spatial order in which auditory stimuli were presented. The “left-leading” group ( $N = 11$ ; 6 women,

5 men) were presented with lead-lag auditory stimuli (conditions  $A_{PE}$ ,  $A_{PE}V_{Lead}$ , and D, as defined below) in a left-then-right pattern, while the “right-leading” group ( $N = 10$ ; 6 women, 4 men) were presented with lead-lag auditory pairs in a right-then-left pattern. Single noise bursts (condition A, as defined below) were presented from the leading loudspeaker (i.e., right loudspeaker for right-leading and left loudspeaker for left-leading stimuli).

**Experimental setup.** The experimental setup used here is described in a recent report (Bishop et al. 2011) and is depicted graphically in Fig. 1A. Briefly, participants sat in a comfortable chair with their chins placed in a chinrest  $\sim 120$  cm from a computer monitor (Dell FPW2407) positioned between two free-field loudspeakers (Tannoy Precision 6) located  $\pm 18^\circ$  to the left and right of the midsagittal plane. Visual stimulation was provided via two white light-emitting diodes (LEDs,  $1.7^\circ$  visual angle, 7 cd) positioned directly above each loudspeaker cone and placed in partially transparent ping-pong balls to generate diffuse light flashes. A computer keyboard was used to collect subject responses.

**Stimuli, trial structure, and task.** Auditory stimuli consisted of 15-ms noise bursts presented at  $\sim 68$  dB(A) from each loudspeaker. Loudspeaker output was calibrated prior to data collection for each participant with a handheld digital sound pressure level (SPL) meter ([www.radioshack.com](http://www.radioshack.com)). Unique noise samples were created for each

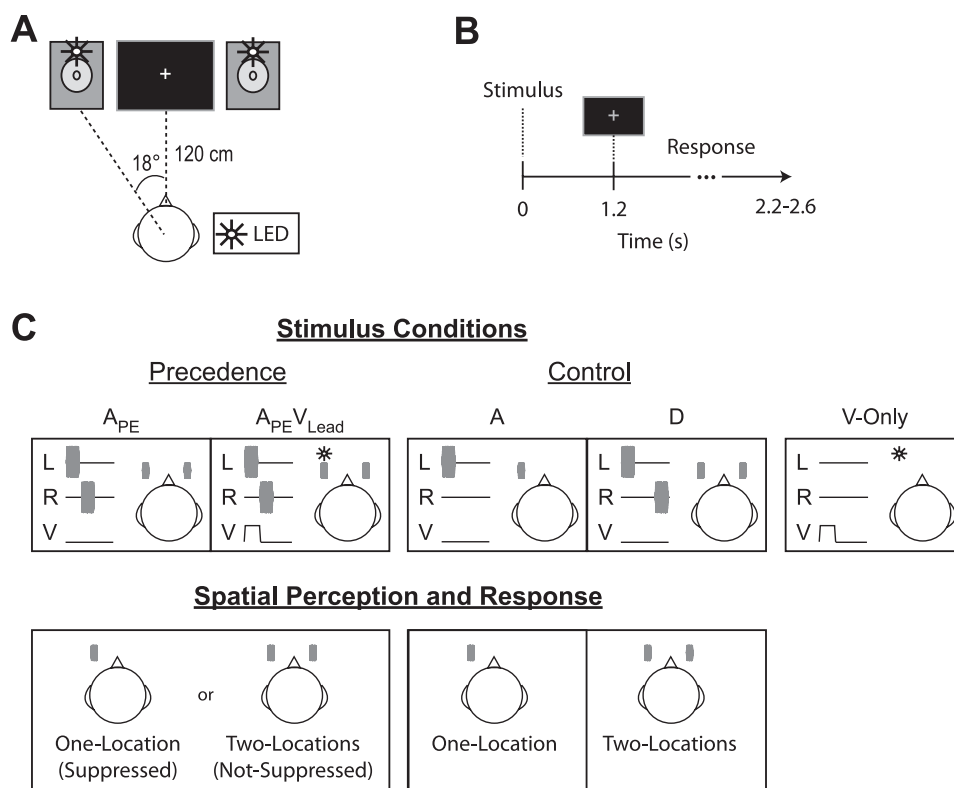


Fig. 1. Setup, stimuli, and trial structure during the EEG session. **A:** listeners sat with their chin placed in a chinrest  $\sim 120$  cm from a computer monitor. The computer monitor was used to present the fixation cross and prompt listeners to respond by changing the color of the cross from white to green. Auditory stimuli (15-ms noise bursts) were presented via 2 free-field loudspeakers located on either side of the computer monitor ( $\sim 18^\circ$  to left and right of midsagittal plane). Visual stimulation was provided via 2 white light-emitting diodes (LEDs, white circles) located above each loudspeaker cone. **B:** each trial began with stimulus presentation ( $<25$  ms), followed by a 1.2-s delay prior to prompting subjects to respond by briefly changing the fixation cross from white to green (shown as gray here for illustration purposes). Subjects had 1.0–1.4 s to respond after the response cue. Trial onsets were jittered from 2.2 to 2.6 s. **C:** stimuli, auditory spatial percepts, and their corresponding response categories are depicted for the left-leading group. The precedence effect (i.e., echo suppression) was elicited by presenting lead-lag auditory pairs unimodally ( $A_{PE}$ ) or paired with a visual flash at the same time and place as the leading sound ( $A_{PE}V_{Lead}$ ). Listeners perceived a sound either from both speakers and responded “two-locations” (i.e., Not-Suppressed) or near the leading speaker and responded “one-location” (i.e., Suppressed). Additional control conditions were included to ensure that subjects were performing the task. These included a single auditory noise burst presented at the leading location (A) and a lead-lag pair with an extended temporal delay (20 ms) to ensure that subjects never suppressed the echo (D). Attentive listeners should respond one-location to all stimuli in the A condition and two-locations to all stimuli in the D condition. Finally, listeners were presented with a visual flash at the leading location in the absence of auditory stimuli (V-only). Listeners were instructed to make no response to the V-only condition. L, left loudspeaker; R, right loudspeaker; V, visual flash.

trial in MATLAB with in-house scripts. Auditory stimuli were presented from a single loudspeaker or one from each loudspeaker with a fixed temporal offset referred to as the lead-lag delay. In other trials, auditory stimuli were paired with visual stimulation provided via 15-ms diffuse light flashes from LEDs positioned above the two loudspeaker cones. An oscilloscope (Tektronix TDS 2004) was used to ensure precise temporal coincidence of audiovisual stimuli and event triggers sent to the EEG acquisition computer. Timing errors rarely exceeded 25  $\mu$ s and never exceeded 0.2 ms.

Subjects were instructed to fixate on a small white cross located at the center of the computer monitor. Trials began with stimulus presentation (dependent on the subject-specific lead-lag delay, but generally <25 ms), and subjects were cued to respond 1.2 s after stimulus onset by briefly changing the white fixation cross to green (Fig. 1B). This response delay was imposed to emphasize response accuracy over shorter reaction times. Subjects answered two questions via two sequential button presses following the response cue. First, subjects responded whether they heard sound from the left of midline (left-side), the right of midline (right-side), or both sides (both-sides) by pressing their right index, ring, and middle fingers, respectively. Second, subjects indicated whether they heard sound from one location or two locations with their left index and middle fingers, respectively. The response window ranged from 1.0 and 1.4 s (Fig. 1B). Subjects were informed that they had <1 s to enter both responses after the response cue.

**Prescreening procedure.** All 39 subjects participated in a 1.5-h behavioral screening session prior to enrollment in the EEG study. The purpose of this session was threefold: 1) to train the subjects on the task, 2) to estimate each subject's echo threshold (i.e., the temporal lead-lag delay at which subjects suppress the temporally lagging sound or "echo" on 50% of trials), and 3) to identify and excuse subjects who were unable to perform the task or had echo thresholds below 2 ms. A 2-ms cutoff was used to ensure that sounds were well lateralized (Bishop et al. 2011) and that the experimenter had a sufficient range for adjusting the lead-lag delay during the EEG experiment (see below). The session began with a verbally narrated PowerPoint presentation to familiarize subjects with the stimuli and task. Subjects were then presented with a battery of stimuli representative of each condition type present in the experiment. These consisted of lead-lag auditory pairs in the absence of visual stimulation ( $A_{PE}$ ), the same lead-lag auditory pair with a contemporaneous visual stimulus at the lead location ( $A_{PE}V_{Lead}$ ), and several controls including a single noise burst presented at the leading location (A), a lead-lag auditory pair with a fixed temporal delay well above subjects' echo thresholds (20- or 100-ms delay, D), and a single noise burst at the lead location paired with a visual stimulus at the lag location ( $AV_{Lag}$ ). Importantly, the leading and lagging noise bursts were physically identical within a given trial for all lead-lag auditory stimulus configurations ( $A_{PE}$ ,  $A_{PE}V_{Lead}$ , and D). Subjects then completed a 5-min practice session with these stimuli to learn the response mapping. After the practice session, subjects completed a short (1–5 min) adaptive procedure to estimate their echo threshold. To do this, subjects were presented with stimuli in only the  $A_{PE}$  configuration with the lead-lag delay initialized to 2 ms and incremented or decremented by 0.5 ms following "one-location" (Suppressed) or "two-locations" (Not-Suppressed) responses, respectively. The algorithm terminated after the direction of change reversed 10 times. The spatial configuration (e.g., right- or left-leading auditory pairs) was pseudorandomized such that each of the two configurations was represented in three of every six trials. The experimenter then presented  $A_{PE}$  stimuli manually to better identify the subject's echo threshold; as in our previous reports (Bishop et al. 2011; London et al. 2012), this often required estimated thresholds to be decreased, likely because of early practice effects. Subjects were then assigned to either the right- or left-leading group of listeners. In the event that subjects had approximately equivalent echo thresholds for each spatial configuration, the experimenter assigned them to a random group. However,

in most cases subjects showed a pronounced asymmetry in their echo thresholds (Clifton and Freyman 1989; Grantham 1996; Saberi et al. 2004). In such cases, subjects were assigned to the leading-side group with the longer of the two echo thresholds.

Once the experimenter assigned a subject to a group, he ran a brief behavioral session that was virtually identical to the EEG study with two exceptions: 1) the sessions contained half the number of trials as an EEG session and 2) an additional control was included. The additional control consisted of a single noise burst presented at the lead location and a simultaneous visual stimulus at the lag location (condition  $AV_{Lag}$ ). This additional control helped to identify and exclude subjects who were susceptible to the ventriloquist's illusion (Howard and Templeton 1966; Recanzone 2009) under the current setup. The data from this behavioral control session were then analyzed, the echo threshold adjusted, and the process repeated until the subject responded "one-location" (i.e., Suppressed) on ~50% of  $A_{PE}$  trials. Finally, the subject completed two additional behavioral blocks similar in duration to the EEG sessions with the estimated echo threshold, but with the addition of the  $AV_{Lag}$  control (20 trials). Subjects were only asked to participate in the EEG experiment if their performance in the control conditions (A, D,  $AV_{Lag}$ ) during these last two blocks was above 80% accuracy and they suppressed the echo on at least 30% of trials in the  $A_{PE}$  configuration. The 30% suppression cutoff was strictly enforced because it was well above the minimal percentage of Suppressed responses needed in the EEG session to ensure enough trials for each condition and percept (~20% or 100 trials). Importantly, a subject's percentage of Suppressed responses in the  $A_{PE}V_{Lead}$  condition had no impact on his/her enrollment in the EEG study.

**EEG recording and procedure.** Sixty-four-channel EEG data were acquired with a BioSemi ActiveTwo system ([www.biosemi.com](http://www.biosemi.com)) with DC amplifiers. Data were low-pass filtered (5th-order sinc filter with 204.8 Hz corner frequency) prior to digitization at 1,024 Hz to prevent aliasing. The EEG experiment lasted ~2.5–3 h including setup time and consisted of at least eight blocks lasting ~7 min each. Each block consisted of 180 trials: 60  $A_{PE}$ , 60  $A_{PE}V_{Lead}$ , 20 obvious singles (A), 20 obvious doubles (20-ms fixed delay, D), and 20 vision only (V-only) (Fig. 1C, top). Importantly, stimuli were presented only as right- or left-leading for each subject. As described in the prescreening session above, subjects were typically assigned to whichever group seemed to have the longer echo threshold. Subjects were provided an optional break approximately every 2 min to ensure that they remained alert during the experiment. The experiment began with the lead-lag delay set to the echo threshold established in the prescreening session. If a subject responded "one-location" (i.e., Suppressed) on <30% or >70% of trials in the  $A_{PE}$  or  $A_{PE}V_{Lead}$  conditions, the experimenter adjusted the lead-lag delay to increase the chances of having an adequate number of trials for each percept in these two conditions. This adjustment was often necessary and in some extreme cases had to be repeated several times before beginning the main experiment. Importantly, data prior to establishing the final lead-lag delay were not included in the analysis. That is, the lead-lag auditory pairs eliciting Suppressed (one-location) and Not-Suppressed (two-locations) responses in the  $A_{PE}$  and  $A_{PE}V_{Lead}$  trials were physically identical, ensuring that any observed differences are strongly linked to genuine perceptual changes and not stimulus differences.

**Event-related potential analysis.** Data were re-referenced off-line to the average of all 64 electrodes in EEGLAB v. 7.2.9.20b. An average reference was used in favor of a mastoid reference for several reasons. First, several previous investigations of echo suppression have utilized an average reference (Backer et al. 2010; Spierer et al. 2009b). Second, many evolving analyses, such as source imaging techniques, rely on measures of topographic variability (e.g., global field power or GFP) that relate most intuitively to an average reference (Lehmann and Skrandies 1980; Murray et al. 2008). Finally, nearly all previous reports from our group have used an average reference (Backer et al. 2010; Kerlin et al. 2010; Shahin et al. 2008,



2010, 2011) and we strove for continuity of presentation. The data were then downsampled to 512 Hz, high-pass filtered in ERPLAB v. 1.0.0.33 (2nd-order Butterworth, 0.5 Hz high-pass cutoff), and decomposed into 64 independent components with EEGLAB's independent components analysis (ICA) (Delorme and Makeig 2004). The ICA decomposition invariably identified a single component with a time course and topography consistent with an eyeblink referred to here as the "eyeblink component." The eyeblink component and any clearly identifiable muscle artifacts were removed from each subject's data prior to low-pass filtering (40-Hz low-pass cutoff, 4th-order Butterworth filter) and epoching (−100 to 500 ms) the data in ERPLAB ([erplab.org](http://erplab.org)). Epochs were flagged for rejection if any sample exceeded  $\pm 80 \mu\text{V}$  at any electrode site during the epoch. Because eyeblinks were removed prior to artifact rejection, it was possible to include trials when subjects had their eyes closed during stimulus presentation in the event-related average. This could introduce a significant confound in the present design. We took several precautions to minimize eyeblinks during stimulus presentation and remove the rare remaining instances. First, subjects were monitored via a remote camera and told explicitly not to blink during stimulus presentation. Although this kept eyeblinks during stimulus presentation to a minimum, it did not completely eliminate such events. Second, we employed an objective means to remove any remaining suspect trials. To do so, the ICA-decomposed data were remixed with only the eyeblink component. This "eyeblink only" data set was then filtered and epoching in an identical fashion but subjected to a much more liberal rejection criterion of  $\pm 30 \mu\text{V}$  applied over the −100 to 100 ms time window. Any trial flagged by either rejection criterion was omitted from the analysis. The remaining trials were binned by condition and response (one-location or two-locations) using ERPLAB to generate condition and response specific event-related potentials (ERPs). Subjects were only included in the analysis if they had >100 trials for each response category (one-location and two-locations) in both the  $A_{PE}$  and  $A_{PE}V_{Lead}$  conditions.

To reduce the data to a more manageable size, the 64-channel electrode montage was divided into three rostrocaudal regions [anterior, central, and posterior (ACP)] and further subdivided sagittally into three sections [left, middle, and right (LMR)] for a total of nine regions (Fig. 2). ERP amplitudes for each region were averaged over five electrodes, indicated as filled black circles in Fig. 2. For each time window analyzed, the mean amplitude of the ERP was included as the dependent measure in an analysis of variance (ANOVA) including

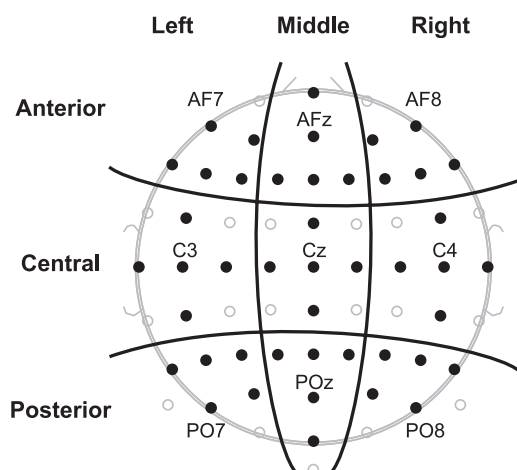


Fig. 2. EEG statistical model. The 64-channel electrode array was divided into 3 rostrocaudal sections (anterior, central, and posterior) and sagittally into 3 additional sections (left, middle, and right) for a total of 9 distinct regions. Mean event-related potential (ERP) amplitude for conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$  and each response, Suppressed or Not-Suppressed, was averaged over 5 electrodes (black circles) within each region. Representative electrodes plotted in subsequent figures are also labeled.

[Condition ( $A_{PE}$  or  $A_{PE}V_{Lead}$ )]  $\times$  [Percept (Suppressed or Not-Suppressed)]  $\times$  [anterior/central/posterior (ACP)]  $\times$  [left/middle/right (LMR)] as within-subject factors and [Leading-Side (right- or left-leading)] as a between-subject grouping factor.

**Behavior analysis.** Although subjects were asked to respond to two different questions for each trial (left-side, right-side, or both-sides and one-location or two-locations), these two questions provided almost completely redundant information. As a result, we used the one-location/two-locations responses to classify trials as Suppressed (S) or Not-Suppressed (NS), respectively, for stimuli in the  $A_{PE}$  and  $A_{PE}V_{Lead}$  configurations. As explained in our previous reports (Bishop et al. 2011; London et al. 2012), this response category measures spatial suppression of the lagging sound or "echo" more accurately. To assess the hypothesized visually enhanced echo suppression, we included the percentage of Suppressed responses in a one-factor [Condition ( $A_{PE}$  or  $A_{PE}V_{Lead}$ )] within-subject ANOVA and included [Leading-Side] as a between-subject grouping factor.

Additionally, we assessed vision's contributions to a phenomenon known as "buildup" of the precedence effect (see Litovsky et al. 1999 for review). The buildup effect manifests with repeated presentation of lead-lag auditory pairs with a fixed lead-lag delay. Specifically, subjects are less likely to report hearing the lagging sound or "echo" after repeated presentation of the same lead-lag pair. This is believed to be a consequence of a listener's ability to build up expectations about the current acoustic scene. We analyzed vision's influence using two novel measures: "Initiation" and "Maintenance." Intuitively, Initiation provides a metric of how quickly buildup is established or "initialized," while Maintenance is a metric of how likely buildup is to persist over time. Initiation is defined here as the probability of a Suppressed response following a Not-Suppressed response in either the  $A_{PE}$  or  $A_{PE}V_{Lead}$  condition. Initiation in the absence and presence of visual information is calculated respectively as

$$\# \{A_{PE}(NS) \rightarrow [A_{PE}(S) \text{ or } A_{PE}V_{Lead}(S)]\} / \# \{A_{PE}(NS) \rightarrow [A_{PE}(S) \text{ or } A_{PE}(NS) \text{ or } A_{PE}V_{Lead}(S) \text{ or } A_{PE}V_{Lead}(NS)]\}$$

and

$$\# \{A_{PE}V_{Lead}(NS) \rightarrow [A_{PE}(S) \text{ or } A_{PE}V_{Lead}(S)]\} / \# \{A_{PE}V_{Lead}(NS) \rightarrow [A_{PE}(S) \text{ or } A_{PE}(NS) \text{ or } A_{PE}V_{Lead}(S) \text{ or } A_{PE}V_{Lead}(NS)]\}$$

In contrast, Maintenance is defined as the probability of a Suppressed response following a Suppressed response in the  $A_{PE}$  or  $A_{PE}V_{Lead}$  condition. Mathematically, Maintenance is the absence and presence of visual information is calculated respectively as

$$\# \{A_{PE}(S) \rightarrow [A_{PE}(S) \text{ or } A_{PE}V_{Lead}(S)]\} / \# \{A_{PE}(S) \rightarrow [A_{PE}(S) \text{ or } A_{PE}(NS) \text{ or } A_{PE}V_{Lead}(S) \text{ or } A_{PE}V_{Lead}(NS)]\}$$

and

$$\# \{A_{PE}V_{Lead}(S) \rightarrow [A_{PE}(S) \text{ or } A_{PE}V_{Lead}(S)]\} / \# \{A_{PE}V_{Lead}(S) \rightarrow [A_{PE}(S) \text{ or } A_{PE}(NS) \text{ or } A_{PE}V_{Lead}(S) \text{ or } A_{PE}V_{Lead}(NS)]\}$$

**Statistical analysis.** All statistical analyses were performed in STATISTICA v. 10 ([www.statsoft.com](http://www.statsoft.com)). Reported *P* values for the ANOVA are corrected for nonsphericity with Greenhouse-Geisser correction where appropriate. Significant effects were further dissected with post hoc *t*-tests. Unless noted otherwise, all summary statistics reflect means  $\pm$  SE.

## RESULTS

**Behavior.** The temporal delay of lead-lag auditory pairs was calibrated for each individual to target 50% one-location responses in the  $A_{PE}$  condition. Importantly, "one-location" and "two-locations" responses to physically identical stimuli in the  $A_{PE}$  and  $A_{PE}V_{Lead}$  conditions were used as a subjective measure of echo suppression. If a subject responded "one-loc-

tion,” the echo was perceptually “Suppressed.” In contrast, if a subject responded “two-locations,” the echo was “Not-Suppressed.” The temporal delay between the two spatially distinct sounds, referred to as an individual’s “echo threshold,” was  $4.89 \pm 0.62$  ms and was statistically indistinguishable between groups by a two-sample *t*-test [ $5.30 \pm 0.875$  ms for left-leading and  $4.45 \pm 0.918$  ms for right-leading,  $F_{(1,19)} = 0.444$ ,  $P = 0.513$ ; mean  $\pm$  SE]. This calibration procedure led to  $49.27 \pm 2.86\%$  Suppressed responses in the  $A_{PE}$  condition ( $51.39 \pm 3.95\%$  left-leading,  $47.15 \pm 4.45\%$  right-leading) and  $60.29 \pm 2.56\%$  Suppressed responses in the  $A_{PE}V_{Lead}$  condition ( $59.81 \pm 3.53\%$  left-leading,  $60.78 \pm 3.70\%$  right-leading) (Fig. 3A). The percentage of Suppressed responses was included in an ANOVA with [Condition ( $A_{PE}$  or  $A_{PE}V_{Lead}$ )] as a within-subject factor and [Leading-Side (right- or left-leading)] as a between-subject grouping factor. The main effect of Condition was significant [ $F_{(1,19)} = 27.617$ ,  $P < 0.001$ ], while Leading-Side [ $F_{(1,19)} = 0.107$ ,  $P = 0.75$ ] and the [Condition  $\times$  Leading-Side] interaction [ $F_{(1,19)} = 1.546$ ,  $P = 0.23$ ] were not. This suggests that while the addition of a visual stimulus at the same time and place as the temporally leading noise burst increased the percentage of Suppressed responses by  $11.02 \pm 2.10\%$ , thus replicating our previous findings (Bishop et al. 2011), vision’s contributions to echo suppression did not depend on spatial configuration. Additionally, subjects performed

remarkably well in all control conditions, responding one-location on  $97.35 \pm 0.61\%$  in the obvious single condition (A) and responding one-location on  $0.54 \pm 0.18\%$  (i.e.,  $99.46 \pm 0.18\%$  two-locations responses) in the obvious double (D) condition.

In addition to demonstrating a visual contribution to echo suppression generally, we hypothesized that vision specifically contributes by helping listeners establish or “build up” a set of expectations about the current listening environment. Previous work has shown that echo suppression increases with repeated presentation of lead-lag auditory pairs in the absence of explicit visual stimulation (see Litovsky et al. 1999 for review). That is, subjects are more likely to suppress the echo with repeated presentation. We hypothesized that vision contributes to the buildup process by 1) helping listeners “initialize” expectations more quickly and/or 2) helping to “maintain” a set of established listening expectations longer. We used two novel measures, “Initiation” and “Maintenance,” to investigate vision’s involvement in each of these distinct processes (see METHODS for details). Prior to testing for a vision-specific contribution to echo suppression, we tested whether or not subjects experienced buildup independently of visual input. To do this, we used the Maintenance measure to quantify buildup because it relates most directly to existing definitions. Specifically, if subjects experience buildup, the probability of consecutive

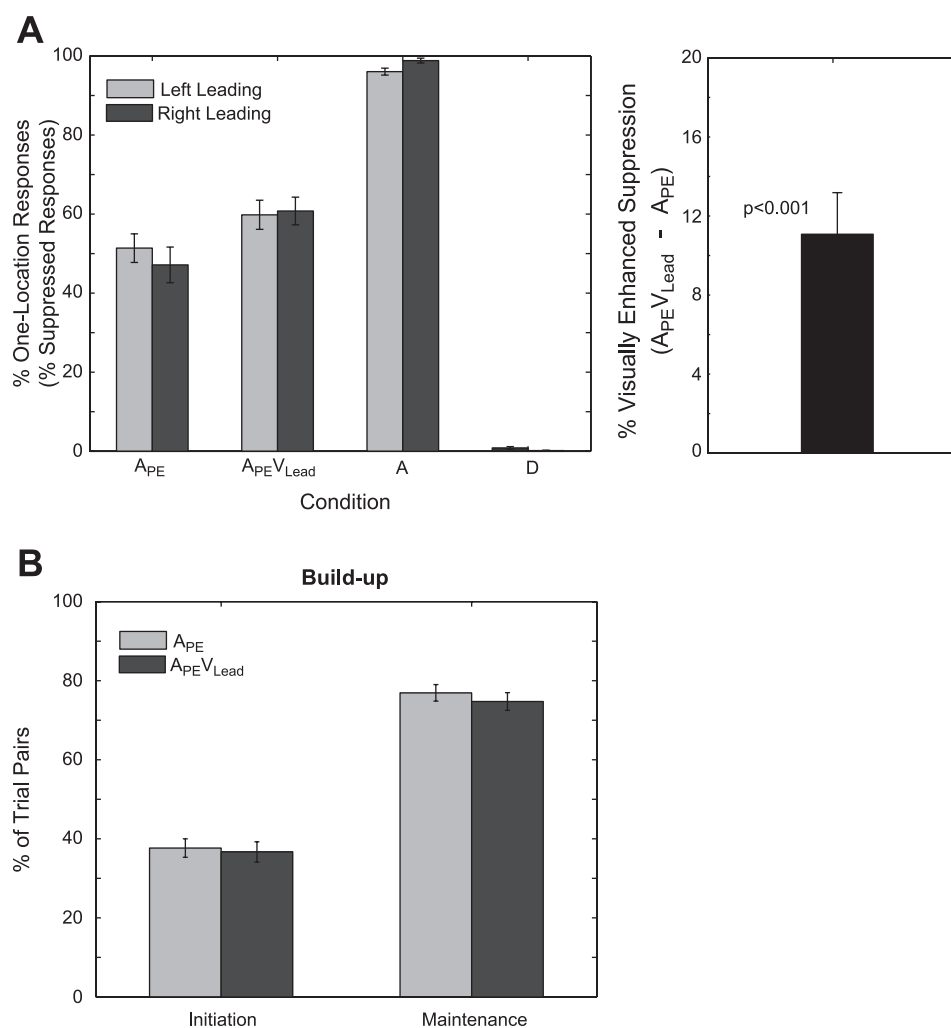


Fig. 3. Subject behavior during the EEG experiment. **A:** the percentage of one-location (Suppressed) responses (left) and the percentage increase in Suppressed responses in condition  $A_{PE}V_{Lead}$  compared with  $A_{PE}$  (right) are plotted for each condition separately for left-leading and right-leading groups. The reader will recall that the percentage of one-location responses and the percentage of Suppressed responses are synonymous, only in the  $A_{PE}$  and  $A_{PE}V_{Lead}$  conditions. The addition of a contemporaneous visual cue at the location of the leading sound (i.e., condition  $A_{PE}V_{Lead}$ ) led to a  $11.02 \pm 2.10\%$  ( $P < 0.001$ ) enhancement of echo suppression. **B:** 2 metrics for buildup, Initiation and Maintenance, are reported for conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$ . Initiation reflects the percentage of  $A_{PE}/A_{PE}V_{Lead}$  trials eliciting a “Suppressed response” following a “Not-Suppressed response” in an immediately preceding  $A_{PE}$  (light gray) or  $A_{PE}V_{Lead}$  (dark gray) trial. In contrast, Maintenance reflects the percentage of  $A_{PE}/A_{PE}V_{Lead}$  trials eliciting a “Suppressed response” following a “Suppressed response” on an immediately preceding  $A_{PE}$  (light gray) or  $A_{PE}V_{Lead}$  (dark gray) trial. Intuitively, Initiation provides a measure for how quickly subjects establish a set of listening expectations, while Maintenance reflects how likely an established set of expectations are to persist over trials. There were no significant differences of either buildup metric between conditions, suggesting that vision does not contribute directly to listener expectations in the context of “buildup.”  $N = 21$  (11 left-leading, 10 right-leading). Error bars reflect SE.

Suppressed responses should be greater than chance. We refer the reader to METHODS for a more formal, mathematical treatment. Maintenance was  $76.92 \pm 2.10\%$  after a Suppressed response to an  $A_{PE}$  stimulus and  $74.73 \pm 2.23\%$  after a Suppressed response to an  $A_{PE}V_{Lead}$  stimulus (Fig. 3B). A one-sample  $t$ -test demonstrated that Maintenance was significantly different from chance (50%) across both conditions [ $F_{(1,20)} = 156.8$ ,  $P < 0.001$ ]. Despite clear evidence that listeners experienced buildup, there were no significant differences in Maintenance in the presence or absence of visual information—that is, after a Suppressed response on an  $A_{PE}V_{Lead}$  or  $A_{PE}$  trial, respectively [ $F_{(1,20)} = 2.7312$ ,  $P = 0.11$ ]. In addition to Maintenance, we used a second measure of buildup, “Initiation,” to test whether visual stimulation facilitates the initial establishment of a listener’s expectations. Initiation was  $37.68 \pm 2.10\%$  after a Not-Suppressed response to an  $A_{PE}$  stimulus and  $36.68 \pm 2.58\%$  after a Not-Suppressed response to an  $A_{PE}V_{Lead}$  stimulus. A paired  $t$ -test revealed no

significant differences in Initiation between the two conditions [ $F_{(1,20)} = 0.496$ ,  $P = 0.490$ ]. Put simply, listeners experienced buildup, but there is no evidence that vision contributes to the buildup process directly.

**Sensory evoked potentials.** The ERPs in conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$  recorded at electrode Cz, a site used to describe the stereotypical auditory evoked response, are plotted in Fig. 4A. Auditory lead-lag pairs elicited a classic auditory evoked potential, beginning with the auditory P1 from 30 to 60 ms, followed by the N1 from 70 to 100 ms and P2 from 150 and 180 ms. The auditory P1 is reflected topographically as a positivity over frontal sites, while the N1 and P2 are reflected as negative and positive distributions with a minimum and maximum at Cz, respectively (see Fig. 4B). Please note that despite qualitatively similar ERPs and topographies in conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$ , the  $A_{PE}V_{Lead}$  ERP reflects responses to both auditory and visual stimulation.

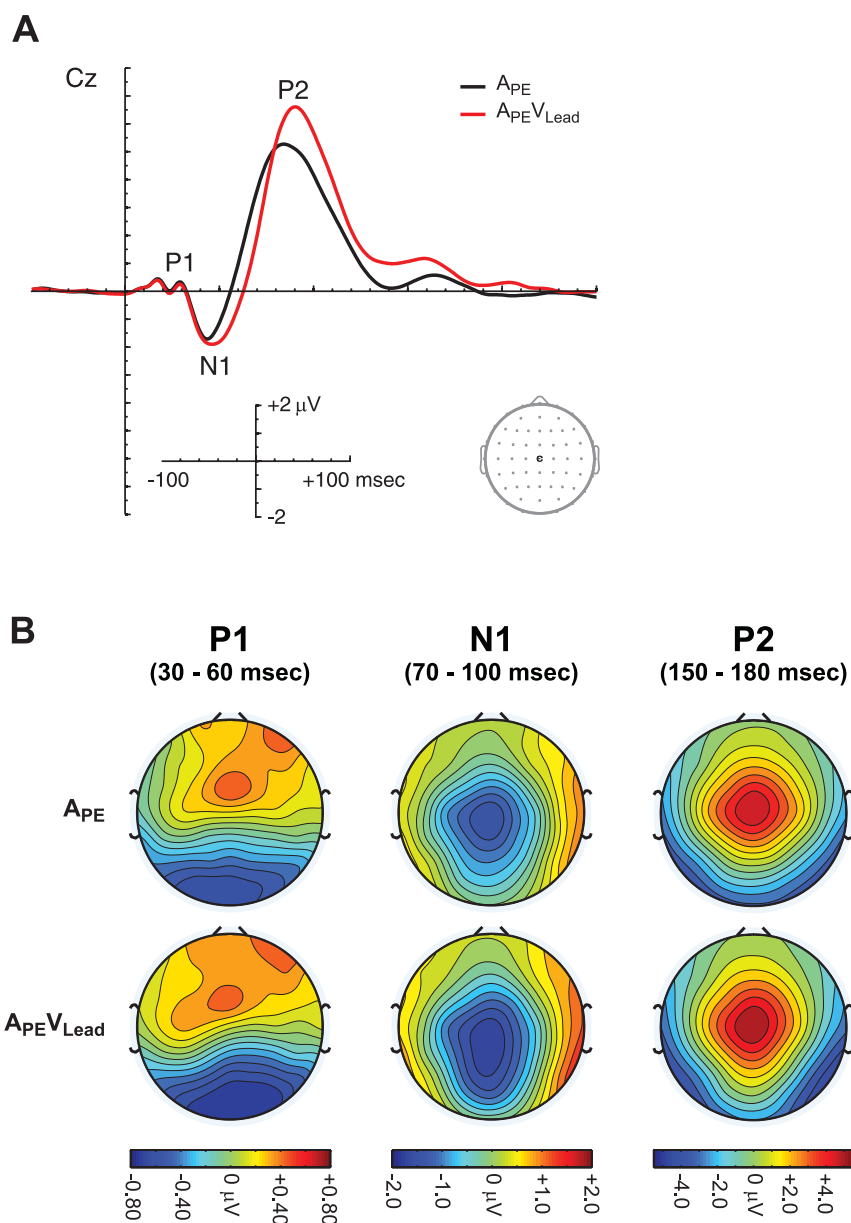


Fig. 4. ERPs to lead-lag auditory pairs. **A**: grand average ERPs for conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$  are plotted at electrode Cz. ERPs comprise the auditory P1, N1, and P2. **B**: scalp topographies during the auditory P1 (30–60 ms), N1 (70–100 ms), and P2 (150–180 ms) are plotted for conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$ . Please note that, although the  $A_{PE}$  and  $A_{PE}V_{Lead}$  topographies are qualitatively similar, the  $A_{PE}V_{Lead}$  ERP and topographies comprise responses to both auditory and visual stimuli.

**Time course of echo suppression.** We began our investigation of the visual enhancement of echo suppression by analyzing the mean ERP amplitude during the auditory P1 (30–60 ms) and N1 (70–100 ms). Specifically, ERPs for conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$  were further binned by response (Suppressed or Not-Suppressed), and the mean amplitude was included in an ANOVA (see METHODS for details). This analysis revealed no significant main effects or interactions of percept during the auditory P1 ( $P > 0.05$  for all effects). However, the same analysis identified a significant [Percept  $\times$  LMR] interaction [ $F_{(2,38)} = 4.156$ ,  $P = 0.028$ ] during the auditory N1. Figure 5 conveys the nature of this interaction. First, the ERPs for Suppressed and Not-Suppressed trials, collapsed across conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$  and leading-side groups, are plotted for two representative electrodes, AF7 and PO8, in Fig. 5, *top left* and *bottom left*. Although these electrodes were selected because the difference topography was maximal over them, there was no (statistically significant) difference among anterior, central, and posterior regions. Second, the topography of the difference between the Suppressed and Not-Suppressed trials from 70 to 100 ms is plotted in Fig. 5, *top right*. Note the dipole from right-posterior to left-anterior sites. Post-hoc  $t$ -tests

revealed a significant difference between Suppressed and Not-Suppressed ERPs over left [ $F_{(1,19)} = 6.341$ ,  $P = 0.021$ ] and right [ $F_{(1,19)} = 5.502$ ,  $P = 0.030$ ], but not middle [ $F_{(1,19)} = 0.0486$ ,  $P = 0.828$ ], electrode sites. Together, these data suggest that general echo suppression mechanisms are engaged during early sensory processing but are unaffected by spatial configuration (e.g., right- or left-leading) or visual input.

In contrast to the early, spatially independent suppression mechanisms engaged during the auditory N1, we find evidence of spatially dependent suppression mechanisms following the auditory N1. The Suppressed and Not-Suppressed waveforms at representative electrodes from anterior-left (AF7), anterior-right (AF8), posterior-left (PO7), and posterior-right (PO8) regions for both right- and left-leading groups are plotted in Fig. 6A. Additionally, the topographies of the Suppressed vs. Not-Suppressed comparison from 100 to 300 ms are depicted in Fig. 6B in 50-ms time bins. Qualitatively, the topographies reflect a rather striking lateralization effect, similar to those seen with visually induced changes in perceived auditory location (Bonath et al. 2007; Donohue et al. 2011). Specifically, the topographical difference maps in Fig. 6B reveal that electrodes ipsilateral to the direction of the perceived spatial

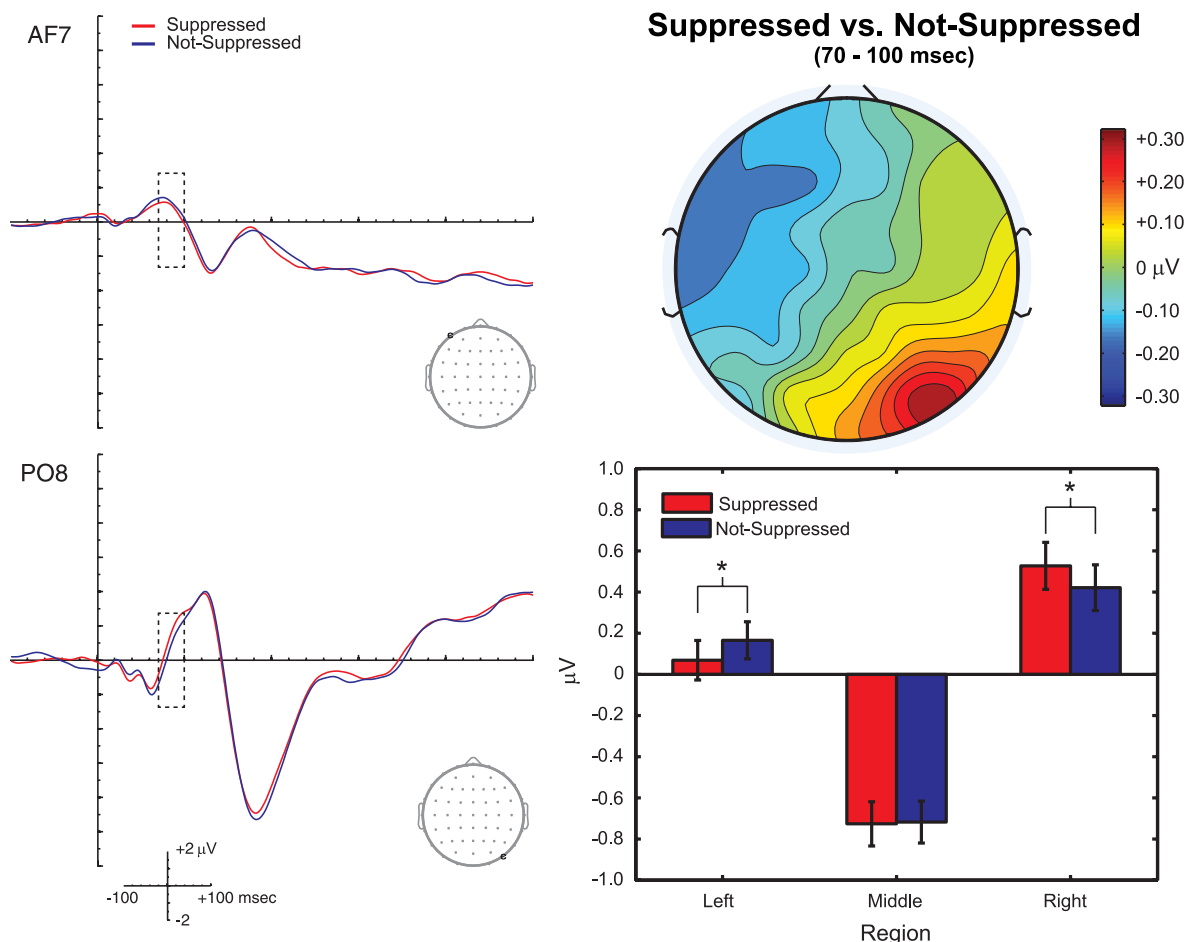
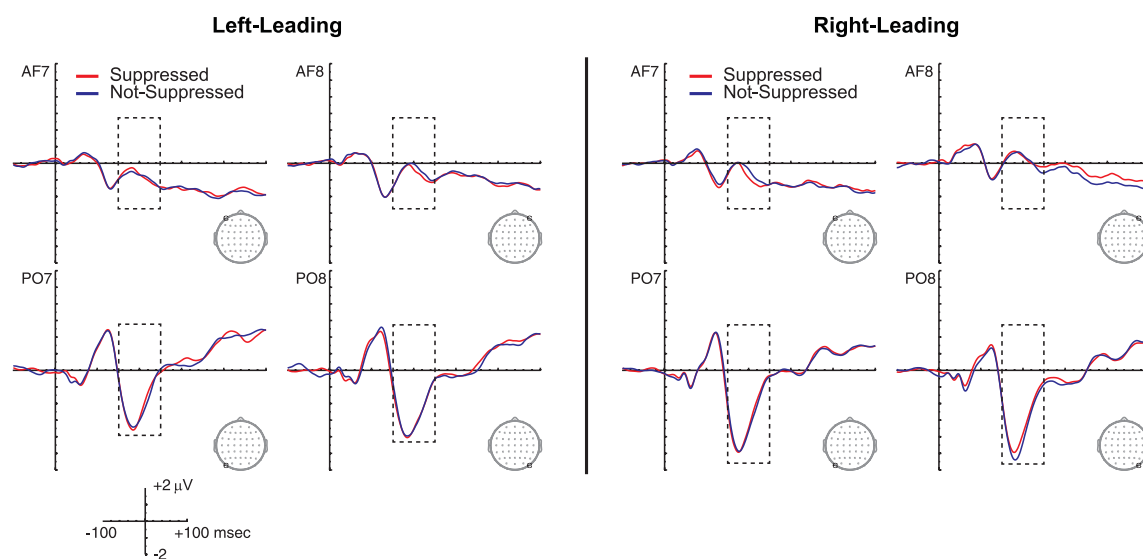
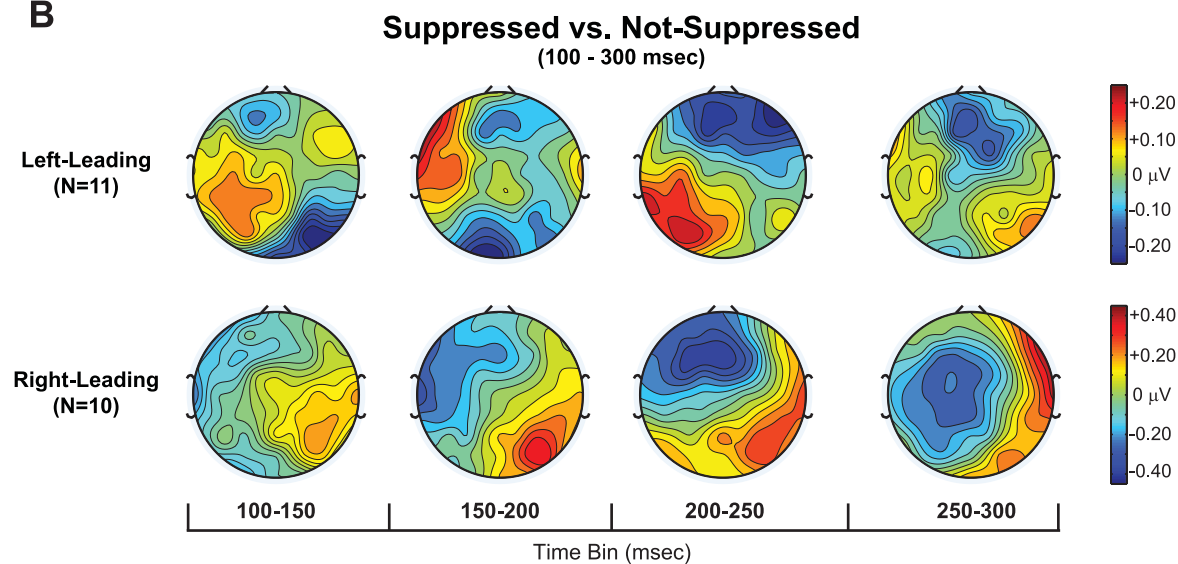
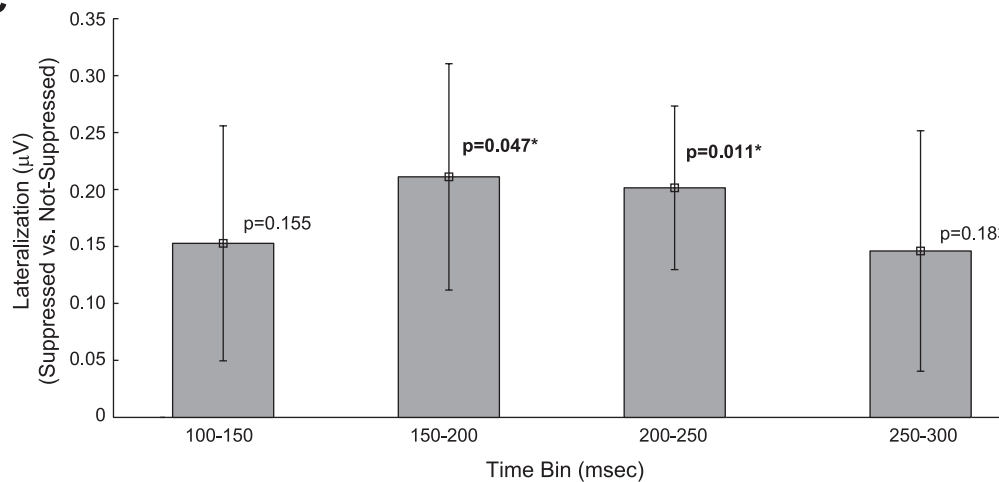


Fig. 5. Perceptual correlates of echo suppression during the auditory N1 (70–100 ms). ERPs for  $A_{PE}$  and  $A_{PE}V_{Lead}$  trials eliciting a Suppressed response or a Not-Suppressed response are plotted for electrodes AF7 (*top left*) and PO8 (*bottom left*). Dashed box indicates the time window of interest. Topography of the Suppressed vs. Not-Suppressed comparison is presented at *top right*, and the mean amplitude for Suppressed responses and Not-Suppressed responses is plotted at *bottom right* for left, middle, and right electrode sites. Perceptual differences were found over left ( $P < 0.05$ ) and right ( $P < 0.05$ ), but not middle ( $P > 0.05$ ), sections. Please note that although AF7 and PO8 differ in their location along the anterior to posterior axis, there were no statistically significant differences along the anterior, central, and posterior (ACP) regions. Instead, these electrodes were selected to complement the graphical summary in the bar plot.  $N = 21$ . Error bars reflect SE. \* $P < 0.05$ .

**A****B****C**



shift of the echo (e.g., left electrodes for the left-leading group) tend to be more positive while contralateral electrodes tend to be negative in the Suppressed vs. Not-Suppressed comparison. Figure 6C plots the Suppressed (ipsi vs. contra) vs. Not-Suppressed (ipsi vs. contra) comparison. Intuitively, this measure describes differential lateralization for Suppressed compared to Not-Suppressed trials. In agreement with our informal observations, an ANOVA during the 200–250 ms time bin revealed a significant [Percept  $\times$  LMR  $\times$  leading-side] interaction [ $F_{(2,38)} = 3.514$ ,  $P = 0.043$ ]. Post hoc tests revealed that this was due to a significant difference in the Suppressed (ipsi vs. contra) vs. Not-Suppressed (ipsi vs. contra) comparison [ $F_{(1,19)} = 7.880$ ,  $P = 0.011$ ]. Although the ANOVAs for the 100–150, 150–200, and 300–350 ms time bins did not yield a similar [Percept  $\times$  LMR  $\times$  Leading-Side] interaction ( $P > 0.05$ ), a targeted test of the hypothesized lateralization effect found in the 200–250 ms time bin revealed a significant effect during the 150–200 ms time bin [ $F_{(1,19)} = 4.512$ ,  $P = 0.047$ ], suggesting that the differential lateralization begins earlier than 200 ms. In contrast, the same comparison was statistically insignificant in the 100–150 [ $F_{(1,19)} = 2.193$ ,  $P = 0.155$ ] and 250–300 [ $F_{(1,19)} = 0.440$ ,  $P = 0.183$ ] ms time bins, suggesting that this lateralization effect is a transient process manifesting from 150 from 250 ms. Importantly, we did not find any evidence of visual contributions in any of these time windows, although there was a trend-level [Condition  $\times$  Percept  $\times$  LMR] interaction in the 200–250 ms time bin [ $F_{(2,38)} = 2.721$ ,  $P = 0.085$ ].

Visual contributions immediately follow the spatially dependent mechanisms described in the previous paragraph. The ERPs for conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$  for Suppressed and Not-Suppressed percepts are plotted for three representative electrodes from anterior-middle (AFz), central-middle (Cz), and posterior-middle (POz) regions in Fig. 7A. ANOVA of the 250–300 ms time bin revealed a significant [Condition  $\times$  Percept  $\times$  ACP  $\times$  Leading-Side] interaction [ $F_{(2,38)} = 4.855$ ,  $P = 0.032$ ]. Post hoc tests of the left-leading group revealed a significant [Condition  $\times$  Percept] interaction over posterior [ $F_{(1,19)} = 5.143$ ,  $P = 0.035$ ] sites, as well as a trend-level effect over anterior electrodes [ $F_{(1,19)} = 4.367$ ,  $P = 0.0503$ ]. The [Condition  $\times$  Percept] interaction over central sites was insignificant for the left-leading group [ $F_{(1,19)} = 0.092$ ,  $P = 0.77$ ]. However, there was no statistically significant [Condition  $\times$  Percept] interaction for the right-leading group within any one of the ACP regions (Fig. 7B, right). In contrast to the seemingly space-specific visual contributions from 250 to 300 ms, a similar analysis of the 300–350 ms time window revealed a spatially independent [Condition  $\times$  Percept] interaction [ $F_{(1,19)} = 4.417$ ,  $P = 0.049$ ]. This region-independent [Condition  $\times$  Percept] interaction progressed to a [Condition  $\times$  Percept  $\times$  ACP] interaction from 350 to 400 ms [ $F_{(2,38)} = 4.764$ ,  $P = 0.036$ ]. Post hoc tests revealed that the [Condition  $\times$  Percept] interaction was primarily reflected over posterior sites [ $F_{(1,19)} = 5.752$ ,  $P = 0.027$ ], although there was a trend-level effect over anterior

sections [ $F_{(1,19)} = 4.306$ ,  $P = 0.052$ ] as well. Together, these findings suggest that vision's influence over echo suppression evolves from a spatially dependent to a more generalized mechanism from 250 to 400 ms.

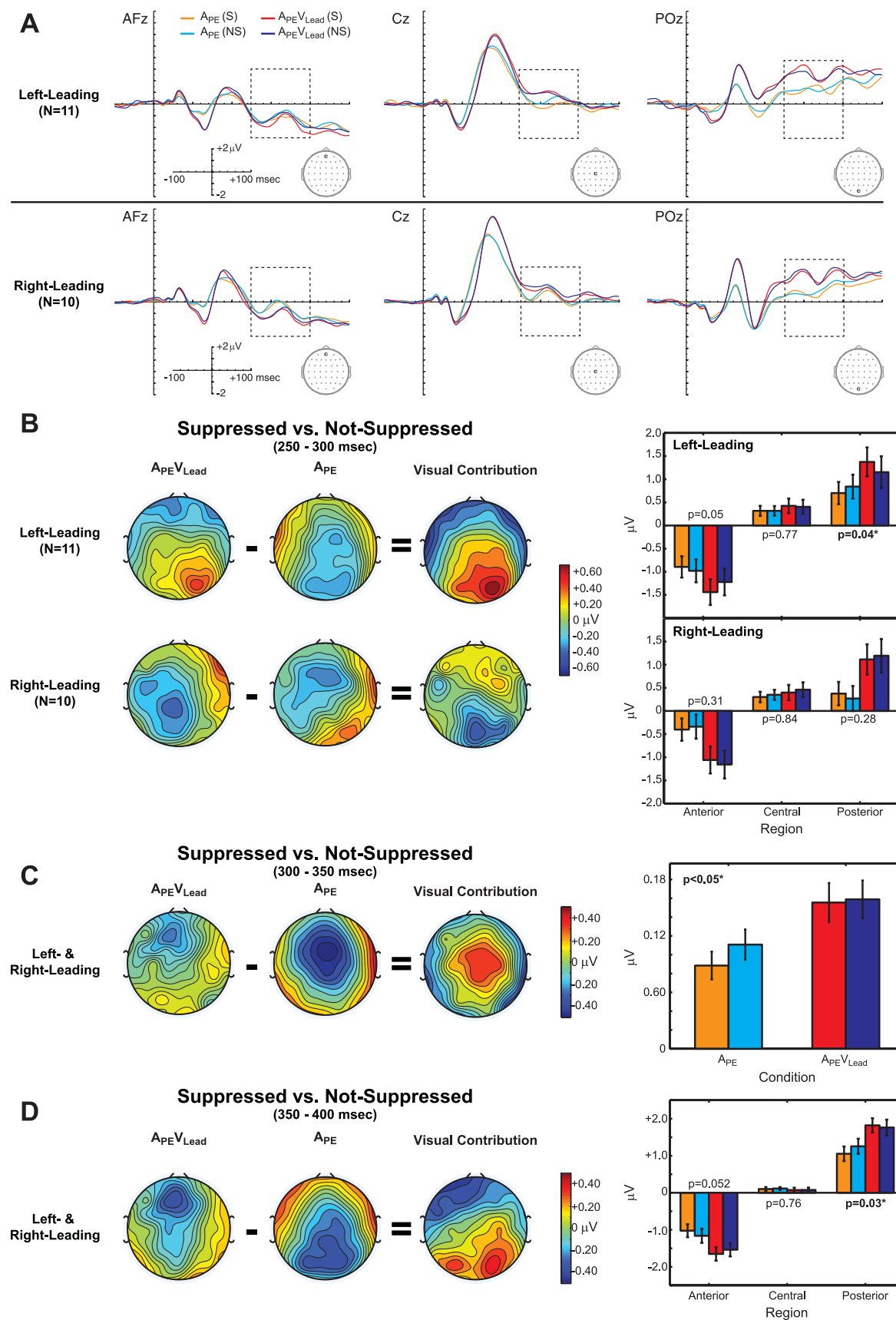
In addition to the mechanisms described above, listeners' spatial perception (i.e., Suppressed or Not-Suppressed responses) is ultimately reflected from 300 to 500 ms. The waveforms for Suppressed and Not-Suppressed trials, collapsed over  $A_{PE}$  and  $A_{PE}V_{Lead}$  conditions, at site CPz are plotted in Fig. 8A. Additionally, the topography of the Suppressed vs. Not-Suppressed comparison is plotted in Fig. 8B and the mean amplitudes for each percept plotted for left, middle, and right sections in Fig. 8C. ANOVA revealed a significant [Percept  $\times$  LMR] interaction [ $F_{(2,38)} = 6.154$ ,  $P = 0.007$ ] during this time window, and post hoc tests revealed a significant difference between Suppressed and Not-Suppressed ERP amplitudes over middle [ $F_{(1,19)} = 17.924$ ,  $P < 0.001$ ], but not left [ $F_{(1,19)} = 0.154$ ,  $P = 0.699$ ] or right [ $F_{(1,19)} = 4.246$ ,  $P = 0.053$ ], sections.

## DISCUSSION

In the present study, we coupled a tightly controlled psychophysical paradigm with 64-channel EEG to elucidate the neural time course of visually enhanced echo suppression in humans. Our findings suggest that echo suppression manifests in several distinct stages, summarized in Fig. 9. Perceptual differences first manifest during the auditory N1. While perception is readily indexed during this critical time period, the mechanisms engaged seem to operate independently of spatial configuration (e.g., right- or left-leading) and visual input. This general suppression mechanism quickly evolves into a spatially lateralized neural response from 150 to 250 ms, likely reflecting the active process of translocating the second sound towards the first. In contrast to the progression from general to spatially specific mechanisms in the absence of visual information, vision first contributes in a space-specific manner from 250 to 300 ms before transitioning to a more generalized contribution from 300 to 400 ms. Ultimately, echo suppression is reflected over middle electrode sites from 300 to 500 ms, a putative index for of varying numbers of attended auditory objects (Alain et al. 2001; Sanders et al. 2008, 2011).

Studies of the precedence effect (i.e., echo suppression) provide compelling evidence that auditory spatial perception remains malleable well into cortical processing, despite strong subcortical representations (Recanzone and Sutter 2008). For instance, although there have been attempts by independent investigators and a host of animal studies implicating the inferior colliculus in echo suppression (Litovsky et al. 1999), no human studies to date have provided evidence of subcortical involvement in echo suppression (Damaschke et al. 2005; Liebenthal and Pratt 1999). In fact, the earliest neural signature of echo suppression in humans is the auditory Pa component (Liebenthal and Pratt 1997, 1999), occurring  $\sim 30$  ms after

Fig. 6. A spatially sensitive correlate of echo suppression manifests from 150 to 250 ms. A: ERP waveforms for Suppressed and Not-Suppressed responses for both  $A_{PE}$  and  $A_{PE}V_{Lead}$  conditions are plotted at representative electrodes at anterior (AFz), central (Cz), and posterior (POz) electrode sites for the left-leading (left) and right-leading (right) groups. Dashed box represents the 100–300 ms time window addressed in the present figure. B and C: topographical difference maps from 100 to 300 ms in 50-ms bins are plotted in B, while the mean differences between the lateralization index (ipsi vs. contra) between Suppressed and Not-Suppressed responses are plotted in C. The lateralization index was significant (\*) from 150 to 250 ms, suggesting a transient, space-specific correlate of echo suppression.  $N = 21$  (11 left-leading, 10 right-leading). Error bars indicate SE.



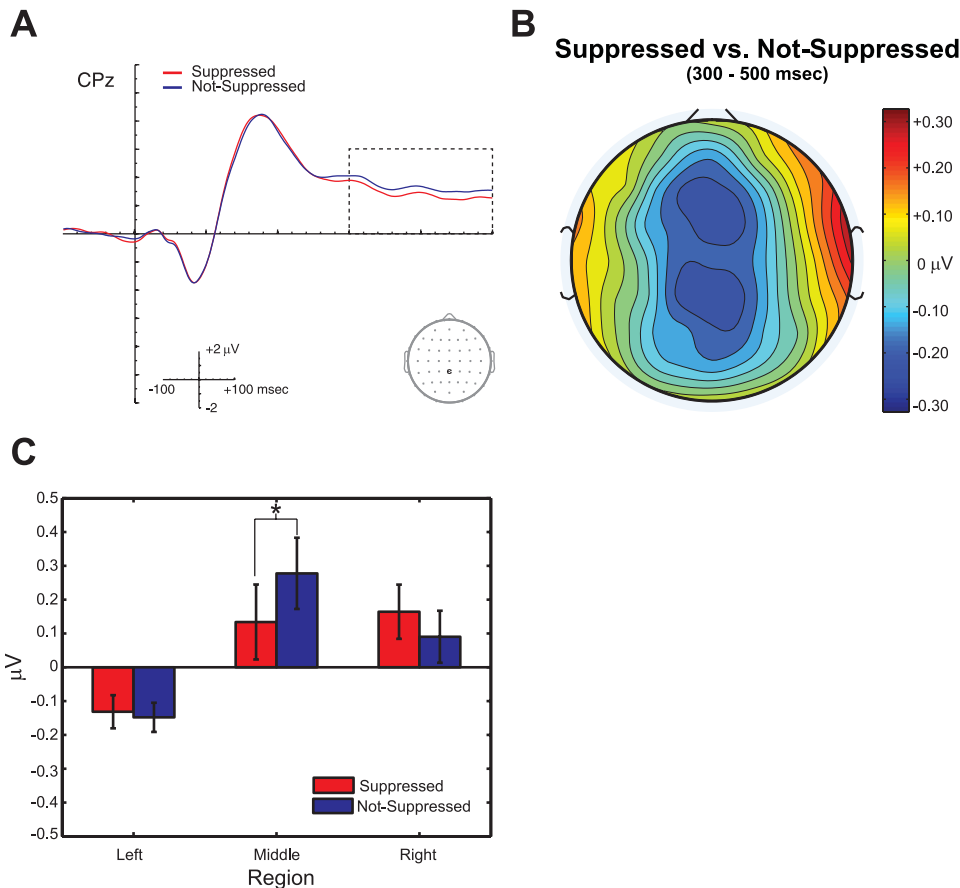


Fig. 8. Perception is reliably indexed from 300 to 500 ms. **A**: Suppressed and Not-Suppressed ERPs, collapsed over conditions  $A_{PE}$  and  $A_{PE}V_{Lead}$ , plotted at electrode CPz. **B**: topographical difference map for the Suppressed vs. Not-Suppressed comparison. **C**: mean amplitude from left, middle, and right sections. Analyses reveal that echo suppression is reflected as a negativity over middle electrode sites from 300 to 500 ms ( $*P < 0.05$ ).  $N = 21$ . Error bars reflect SE.

stimulus onset and thought to reflect early cortical activity (Liebenthal and Pratt 1997). More commonly, human studies, including the present report, suggest that echo suppression has strong neural correlates beginning during or slightly after the auditory N1 that persist for hundreds of milliseconds (Backer et al. 2010; Dimitrijevic and Stapells 2006; Sanders et al. 2008, 2011). The N1 reflects early, albeit not primary, cortical processing, suggesting that echo suppression begins after primary sensory representations have been established. Notably in the present report, the difference topography during the N1 differs from a canonical N1 topography (cf. Figs. 4 and 5), indicating that only a subset of N1 generators are modulated or that an independent process is involved.

Correlates of echo suppression then continue in the moments subsequent to the N1 peak, as observed by other investigators. For instance, Sanders et al. (2008, 2011) have implicated two distinct, putatively cortical correlates of echo suppression. The first is a component over anterior electrode sites from 100 and 250 ms, presumed to be the object-related negativity (ORN) (Alain et al. 2001; Dyson et al. 2005; McDonald and Alain 2005). The authors argue that the ORN reflects early, automatic

detection of multiple auditory objects. Additionally, Sanders et al. report a late, cortical component from 250 to 500 ms over parietal and occipital electrode sites. In contrast to the putative automaticity of the ORN, this late component is thought to index the presence of varying numbers of auditory objects, but only in active listening tasks (Alain et al. 2001). Like Sanders et al., we find distinct correlates of echo suppression from 150 to 250 ms and from 300 and 500 ms. While the latter negativity is virtually identical to that reported by Sanders et al., the mechanisms engaged from 150 to 250 ms are quite different from the ORN in several ways. First, our findings implicate a spatially sensitive component from 150 to 250 ms reminiscent of lateralized responses reported in several audiovisual studies, including one of the ventriloquist's illusion (Bonath et al. 2007) and another on the cross-modal spread of attention (Donohue et al. 2011). Second, the relationship between Suppressed and Not-Suppressed waveforms over anterior sites is reversed during the earlier time window (negative over anterior sites here while positive in Sanders et al. for the Suppressed vs. Not-Suppressed comparison). Despite a similar discrepancy in sign, Sanders et al. (2011) presented subjects with only right-

Fig. 7. Vision contributes to echo suppression from 250 to 400 ms. **A**: ERPs for condition ( $A_{PE}$  and  $A_{PE}V_{Lead}$ ) and percept [Suppressed (S) and Not-Suppressed (NS)] are plotted at representative electrode sites at anterior (AFz), central (Cz), and posterior (POz) electrode sites separately for the left-leading (top) and right-leading (bottom) groups. **B–D**, left: topographical difference maps for the Suppressed vs. Not-Suppressed comparison are shown for each condition from 250 to 300 ms (**B**), from 300 to 350 ms (**C**), and from 350 to 400 ms (**D**). The visual contribution to echo suppression was calculated by subtracting the topographical Suppressed vs. Not-Suppressed difference maps [i.e.,  $[A_{PE}V_{Lead}(S) - A_{PE}V_{Lead}(NS)] - [A_{PE}(S) - A_{PE}(NS)]$ ] for each time bin. Right: mean ERP amplitudes for each Condition and Percept at anterior, central, and posterior sections (**B**, **D**) or collapsed across regions (**C**). Our analyses revealed a [Condition  $\times$  Percept] interaction for the left-leading group over posterior electrodes ( $P = 0.04$ ) from 250 to 300 ms, a global [Condition  $\times$  Percept] interaction from 300 to 350 ms, and finally a [Condition  $\times$  Percept] interaction over posterior electrodes from 350 to 400 ms. Orange,  $A_{PE}(S)$ ; cyan,  $A_{PE}(NS)$ ; red,  $A_{PE}V_{Lead}(S)$ ; dark blue,  $A_{PE}V_{Lead}(NS)$ .

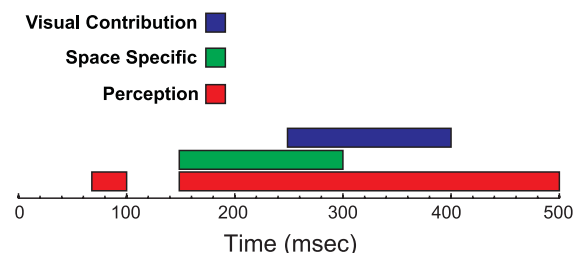


Fig. 9. Schematic of the time course of visual contributions to echo suppression. Time periods that can distinguish between Suppressed and Not-Suppressed percepts are depicted in red, periods of spatial specificity in green, and, finally, visual contributions in blue. The schematic demonstrates that echo suppression progresses from general correlates of suppression (70–100 ms), followed by space-specific mechanisms from 150 to 300 ms and ultimately by visual influences from 250 to 300 ms.

leading lead-lag auditory pairs and found a lateralization effect qualitatively similar to that of our right-leading group, although this is not explicitly stated or quantified in their report. Overall, the present report bolsters existing evidence suggesting that the brain's attempt to suppress distracting echoes and auditory spatial perception generally is a progressive process lasting hundreds of milliseconds. The progressive nature of echo suppression provides vision with ample opportunity to modify the ultimate percept, despite the relatively sluggish visual arrival times in cortex.

Vision's contributions to echo suppression support growing evidence of pervasive audiovisual interactions in cortex. Until recently, audition and vision were assumed to be well insulated in early sensory regions, only converging in higher-level association areas. However, recent evidence from human neuroimaging studies and invasive recordings in animals have demonstrated audiovisual interactions at virtually every level of cortical processing (Driver and Noesselt 2008; Driver and Spence 2000; Ghazanfar and Schroeder 2006; Macaluso 2006). This includes functional visual contributions in early auditory cortex (Allman et al. 2008; Bizley and King 2009; Schroeder et al. 2003), auditory contributions to visual processing (Iurilli et al. 2012; Meredith et al. 2009), and anatomical projections between low-level sensory areas (Clemons et al. 2008; Falchier et al. 2002, 2010; Smiley and Falchier 2009), leading many to conclude that the brain is fundamentally multisensory (Ghazanfar and Schroeder 2006). Here, we find further evidence of the pervasive nature of audiovisual interactions. Although we initially hypothesized that vision contributes to echo suppression by modulating early auditory representations (e.g., during the auditory P1 or N1), we find little supporting evidence in the present study. Instead, visual contributions are realized from 250 to 400 ms, well after sensory representations have been established. Moreover, vision's contribution in the first part of that period (250–300 ms) is only found in the left-leading group. This nuance may be a consequence of hemispheric specialization of acoustic spatial processing (Krumbholz et al. 2005), including echo suppression (Spierer et al. 2009a), or of multisensory integration and attention (Fairhall and Macaluso 2009). Additional studies are necessary to test these hypotheses directly. Although far from definitive, vision's surprisingly late contribution to echo suppression (i.e., largely during late perceptual correlates from 300 to 500 ms) suggests that vision contributes primarily during decision making. Alternatively, vision may contribute by reshaping existing

auditory representations in space-sensitive auditory cortex, such as planum temporale (PT) (Deouell et al. 2007; Lewald et al. 2008; Pavani et al. 2002; Zatorre et al. 2002), or multisensory parietal cortex (Bishop and Miller 2009; Bushara et al. 1999; Macaluso et al. 2004; Miller and D'Esposito 2005). Finally, the visual flash may have served as a cross-modal attentional cue that directs a listener's attention away from the echo's location and thus reduces detectability (Spence and Driver 2004). However, we consider this unlikely for two key reasons. First, existing studies suggest that spatial attention, including reflexive exogenous cuing, requires tens of milliseconds at a minimum for listeners to adequately orient their attention (Warner et al. 1990). Since all our audiovisual onsets occurred within a very narrow temporal window—typically <5 ms—it is unlikely that the visual flash oriented a listener's attention prior to the delivery of the echo itself. Second, the auditory N1 has long been observed to be modulated by attentional demands (Picton and Hillyard 1974). However, we do not find any visual contributions to echo suppression during the N1 time window, suggesting that the N1 is not modulated by visually guided attention. Regardless of the precise neural mechanism, these data clearly demonstrate that vision contributes to echo suppression well after initial sensory processing in cortex.

Several existing reports provide supporting evidence that vision's influence over auditory spatial perception occurs late in cortical processing (Bonath et al. 2007; Donohue et al. 2011). First, a study by Bonath et al. (2007) employed the ventriloquist's illusion to induce a change in the perceived location of a centrally presented auditory stimulus. The authors reported that a visually induced change in auditory perception manifests as a negativity contralateral to the direction of the illusory spatial shift (e.g., over right electrodes for a leftward shift) from 230 to 270 ms. Furthermore, source imaging attempts and a parallel functional magnetic resonance imaging (fMRI) study provided strong, converging evidence linking the lateralized scalp topography to the PT, a spatially sensitive cortical region thought to be functionally homologous to caudomedial and caudolateral belt regions in macaque monkeys (Miller and Recanzone 2009; Recanzone et al. 2000). Interestingly, the lateralized effect from 230 to 270 ms reported by Bonath et al. (2007) is remarkably similar to the lateralized neural response reported here from 150 to 250 ms; this suggests a common time window for both auditorily and visually induced illusions in auditory spatial perception. Second, a study from Donohue et al. found a similar lateralized neural correlate from 200 to 250 ms in response to a presumed visually induced change in perceived auditory space (Donohue et al. 2011). Additionally, Donohue et al. demonstrated that this putative index of spatial perception can be amplified further through cross-modal spatial attention. Together, our findings join a growing body of evidence suggesting that vision's influence over perceived auditory space, be it through cross-modal integration or attentional mechanisms, is not necessarily linked to early auditory representations. Instead, vision is often a late contributor to auditory spatial perception. Additionally, this evidence suggests a critical time window for auditory spatial perception from 200 to 250 ms that can be modulated through auditory (e.g., the precedence effect) and visual (e.g., the ventriloquist's illusion) spatial illusions as well as cross-modal attentional demands. However, we do not



suggest that vision can only contribute to auditory spatial processing through these late, cortical mechanisms. In fact, audiovisual spatial interactions have been demonstrated numerous times in subcortical structures (Keuroghlian and Knudsen 2007; Meredith and Stein 1983; Stein et al. 2004, 2009). Instead, we argue that audiovisual interactions can be found at virtually every level, including late sensory and decision making processes. Despite a relatively late contribution in the present report, vision may make earlier contributions to echo suppression under different circumstances. For instance, a more complex visual stimulus, such as a speaker's mouth movements, may have the added benefit of providing expectations about the arrival time of the leading sound that could further enhance echo suppression. Alternatively, a virtually identical paradigm employing a higher-contrast visual stimulus (e.g., a checkerboard) might reveal an earlier contribution that has gone undetected here with a spatially diffuse flash of light.

Our between-subject design (right- or left-leading groups), although necessary to ensure enough trials for analysis, does impose some interpretational constraints. Specifically, each listener was presented with lead-lag auditory pairs in one spatial configuration (e.g., right-then-left for right-leading subjects). Consequently, top-down influences such as expectations and spatial attention likely contributed to echo suppression. However, we would argue that the involvement of these top-down mechanisms does not limit conclusions regarding echo suppression as they undoubtedly play a central role in spatial hearing in everyday life. For instance, the ventriloquism aftereffect demonstrates that expectations about a sound's location, in this case provided by spatially disparate visual stimuli, can have a profound and lasting impact on auditory spatial perception (Frissen et al. 2005; Kopco et al. 2009; Lewald 2002; Radeau and Bertelson 1977; Recanzone 1998; Woods and Recanzone 2004). In addition, trial-to-trial changes in spatial attention may impact echo suppression. That is, subjects may be more sensitive to the presence of an echo, and thus more likely to respond "two-locations" or "Not-Suppressed," if they are attending to the lag location. This possibility is supported by a recent study demonstrating that simply orienting a listener's spatial attention toward or away from the location of the temporally lagging sound via an exogenous or endogenous attentional cue can have a dramatic impact on echo suppression (London et al. 2012). Echo suppression almost certainly relies on bottom-up (e.g., relative timing) and top-down (e.g., expectations and attention) processes, and further experimentation is necessary to dissociate the contributions of each of these at a neural level.

In addition to strong links between echo suppression and classic ERP measures—such as amplitude and latency—Backer et al. (2010) report enhanced neural precision (i.e., intertrial phase coherence or ITPC) in the alpha/beta and gamma bands during Suppressed vs. Not-Suppressed trials. These analyses are not reported here, pending further investigation. Specifically, we conducted analyses similar to those in Backer et al. (2010) but failed to identify differences between the Suppressed and Not-Suppressed trials in the auditory-only ( $A_{PE}$ ) condition (data not shown). There are several possible explanations for this discrepancy. First, Backer et al. explicitly employed buildup trains, while we did not. It is possible that the reported change in neural coherence is due to a change in listener expectation or entrainment of the auditory click train.

Second, the difference could be technical in nature. When comparing phase locking values (PLVs) between different percepts, Backer et al. first converted their PLVs to a Rayleigh's  $Z$  statistic for each response category. This was done in an attempt to correct for a systematic bias of PLVs with varying trial numbers (i.e., small trial numbers lead to larger PLVs). These Rayleigh's  $Z$  statistics were then directly compared rather than the raw PLVs. Using this method, we did find evidence suggesting that vision contributes to echo suppression through enhanced neural coherence in the theta/alpha and gamma bands beginning as early as the auditory P1 and continuing throughout much of the epoch (data not shown). However, we included additional control analyses that strongly suggested that this procedure is biased by differences in trial numbers that could predict the observed effects. We are currently investigating whether this contributed to the findings reported by Backer et al. (2010).

In sum, the findings reported here contribute to our understanding of auditory spatial perception and audiovisual interactions in humans in several key ways. First, they corroborate previous studies suggesting that echo suppression and, by extension, auditory spatial perception is a slow, progressive process supported by a complex cortical network. Second, these findings suggest that audiovisual interactions generally are not limited to early processing stages but are instead ubiquitous, even in late stages of cortical processing. Finally, we propose that 200–250 ms after stimulus onset is a critical time range for modulating auditory spatial perception, be it through auditory (e.g., the precedence effect) or visual (e.g., ventriloquism) spatial illusions or cross-modal attention. Future studies of echo suppression may focus on dissecting the precise nature of each dissociable stage of echo suppression and the contributions from bottom-up and top-down factors with similar methods or investigate the underlying neural substrates by employing techniques like fMRI. Alternatively, future studies may investigate whether stochastic changes in auditory spatial perception in the absence of an illusory manipulation are linked to the proposed critical time window of 200–250 ms. These advances will undoubtedly provide key insights into our understanding of communication and navigation in noisy environments.

#### ACKNOWLEDGMENTS

We thank Dr. Marty Woldorff for his invaluable insight and helpful discussions.

#### GRANTS

This research was supported by R01-DC-008171 and 3R01-DC-008171-04S1 (L. M. Miller) and F31-DC-011429 (C. W. Bishop) awarded by the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

#### DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

#### AUTHOR CONTRIBUTIONS

Author contributions: C.W.B., S.L., and L.M.M. conception and design of research; C.W.B. and S.L. performed experiments; C.W.B. analyzed data;

C.W.B. interpreted results of experiments; C.W.B. prepared figures; C.W.B. drafted manuscript; C.W.B., S.L., and L.M.M. edited and revised manuscript; C.W.B., S.L., and L.M.M. approved final version of manuscript.

## REFERENCES

- Alain C, Arnott SR, Picton TW. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J Exp Psychol Hum Percept Perform* 27: 1072–1089, 2001.
- Allman BL, Keniston LP, Meredith MA. Subthreshold auditory inputs to extrastriate visual neurons are responsive to parametric changes in stimulus quality: sensory-specific versus non-specific coding. *Brain Res* 1242: 95–101, 2008.
- Backer KC, Hill KT, Shahin AJ, Miller LM. Neural time course of echo suppression in humans. *J Neurosci* 30: 1905–1913, 2010.
- Bishop CW, London S, Miller LM. Visual influences on echo suppression. *Curr Biol* 21: 221–225, 2011.
- Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. *J Cogn Neurosci* 21: 1790–1805, 2009.
- Bizley JK, King AJ. Visual influences on ferret auditory cortex. *Hear Res* 258: 55–63, 2009.
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA. Neural basis of the ventriloquist illusion. *Curr Biol* 17: 1697–1703, 2007.
- Bushara KO, Weeks RA, Ishii K, Catalan MJ, Tian B, Rauschecker JP, Hallett M. Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci* 2: 759–766, 1999.
- Clemo HR, Sharma GK, Allman BL, Meredith MA. Auditory projections to extrastriate visual cortex: connectional basis for multisensory processing in “unimodal” visual neurons. *Exp Brain Res* 191: 37–47, 2008.
- Clifton RK, Freyman RL. Effect of click rate and delay on breakdown of the precedence effect. *Percept Psychophys* 46: 139–145, 1989.
- Cranford J, Diamond IT, Ravizza R, Whitfield LC. Unilateral lesions of the auditory cortex and the “precedence effect.” *J Physiol* 213: 24P–25P, 1971.
- Cranford JL, Oberholtzer M. Role of neocortex in binaural hearing in the cat. II. The “precedence effect” in sound localization. *Brain Res* 111: 225–239, 1976.
- Damaschke J, Riedel H, Kollmeier B. Neural correlates of the precedence effect in auditory evoked potentials. *Hear Res* 205: 157–171, 2005.
- De Gelder B, Bertelson P. Multisensory integration, perception and ecological validity. *Trends Cogn Sci* 7: 460–467, 2003.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134: 9–21, 2004.
- Deouell LY, Heller AS, Malach R, D’Esposito M, Knight RT. Cerebral responses to change in spatial location of unattended sounds. *Neuron* 55: 985–996, 2007.
- Dimitrijevic A, Stapells DR. Human electrophysiological examination of buildup of the precedence effect. *Neuroreport* 17: 1133–1137, 2006.
- Donohue SE, Roberts KC, Grent’t-Jong T, Woldorff MG. The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. *J Neurosci* 31: 7982–7990, 2011.
- Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron* 57: 11–23, 2008.
- Driver J, Spence C. Multisensory perception: beyond modularity and convergence. *Curr Biol* 10: R731–R735, 2000.
- Dyson BJ, Alain C, He Y. Effects of visual attentional load on low-level auditory scene analysis. *Cogn Affect Behav Neurosci* 5: 319–338, 2005.
- Fairhall SL, Macaluso E. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur J Neurosci* 29: 1247–1257, 2009.
- Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22: 5749–5759, 2002.
- Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF. Projection from visual areas V2 and prostriate to caudal auditory cortex in the monkey. *Cereb Cortex* 20: 1529–1538, 2010.
- Frissen I, Vroomen J, de Gelder B, Bertelson P. The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychol (Amst)* 118: 93–100, 2005.
- Ghazanfar AA, Chandrasekaran CF. Paving the way forward: integrating the senses through phase-resetting of cortical oscillations. *Neuron* 53: 162–164, 2007.
- Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? *Trends Cogn Sci* 10: 278–285, 2006.
- Grantham DW. Left-right asymmetry in the buildup of echo suppression in normal-hearing adults. *J Acoust Soc Am* 99: 1118–1123, 1996.
- Howard IP, Templeton WB. *Human Spatial Orientation*. New York: Wiley, 1966.
- Iurilli G, Ghezzi D, Olcese U, Lassi G, Nazzaro C, Tonini R, Tucci V, Benfenati F, Medini P. Sound-driven synaptic inhibition in primary visual cortex. *Neuron* 73: 814–828, 2012.
- Kerlin JR, Shahin AJ, Miller LM. Attentional gain control of ongoing cortical speech representations in a “cocktail party.” *J Neurosci* 30: 620–628, 2010.
- Keuroghlian AS, Knudsen EI. Adaptive auditory plasticity in developing and adult animals. *Prog Neurobiol* 82: 109–121, 2007.
- Kopco N, Lin IF, Shinn-Cunningham BG, Groh JM. Reference frame of the ventriloquism aftereffect. *J Neurosci* 29: 13809–13814, 2009.
- Krumbholz K, Schonwiesner M, von Cramon DY, Rubsamen R, Shah NJ, Zilles K, Fink GR. Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cereb Cortex* 15: 317–324, 2005.
- Lehmann D, Skrandies W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol* 48: 609–621, 1980.
- Lewald J. Rapid adaptation to auditory-visual spatial disparity. *Learn Mem* 9: 268–278, 2002.
- Lewald J, Riederer KA, Lentz T, Meister IG. Processing of sound location in human cortex. *Eur J Neurosci* 27: 1261–1270, 2008.
- Li L, Qi JG, He Y, Alain C, Schneider BA. Attribute capture in the precedence effect for long-duration noise sounds. *Hear Res* 202: 235–247, 2005.
- Liebenthal E, Pratt H. Evidence for primary auditory cortex involvement in the echo suppression precedence effect: a 3CLT study. *J Basic Clin Physiol Pharmacol* 8: 181–201, 1997.
- Liebenthal E, Pratt H. Human auditory cortex electrophysiological correlates of the precedence effect: binaural echo lateralization suppression. *J Acoust Soc Am* 106: 291–303, 1999.
- Litovsky RY, Colburn HS, Yost WA, Guzman SJ. The precedence effect. *J Acoust Soc Am* 106: 1633–1654, 1999.
- London S, Bishop CW, Miller LM. Spatial attention modulates echo suppression. *J Exp Psychol Hum Percept Perform* 22: 222–223, 2012.
- Macaluso E. Multisensory processing in sensory-specific cortical areas. *Neuroscientist* 12: 327–338, 2006.
- Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage* 21: 725–732, 2004.
- McDonald KL, Alain C. Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *J Acoust Soc Am* 118: 1593–1604, 2005.
- Meredith MA, Allman BL, Keniston LP, Clemo HR. Auditory influences on non-auditory cortices. *Hear Res* 22: 222–223, 2009.
- Meredith MA, Stein BE. Interactions among converging sensory inputs in the superior colliculus. *Science* 221: 389–391, 1983.
- Miller LM, D’Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25: 5884–5893, 2005.
- Miller LM, Recanzone GH. Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc Natl Acad Sci USA* 106: 5931–5935, 2009.
- Murray MM, Brunet D, Michel CM. Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr* 20: 249–264, 2008.
- Pavani F, Macaluso E, Warren JD, Driver J, Griffiths TD. A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Curr Biol* 12: 1584–1590, 2002.
- Picton TW, Hillyard SA. Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr Clin Neurophysiol* 36: 191–199, 1974.
- Radeau M, Bertelson P. Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Percept Psychophys* 22: 222–223, 1977.
- Recanzone GH. Interactions of auditory and visual stimuli in space and time. *Hear Res* 258: 89–99, 2009.
- Recanzone GH. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc Natl Acad Sci USA* 95: 869–875, 1998.

- Recanzone GH, Guard DC, Phan ML, Su TK.** Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol* 83: 2723–2739, 2000.
- Recanzone GH, Sutter ML.** The biological basis of audition. *Annu Rev Psychol* 59: 119–142, 2008.
- Saberi K, Antonio JV, Petrosyan A.** A population study of the precedence effect. *Hear Res* 191: 1–13, 2004.
- Sanders LD, Joh AS, Keen RE, Freyman RL.** One sound or two? Object-related negativity indexes echo perception. *Percept Psychophys* 70: 1558–1570, 2008.
- Sanders LD, Zobel BH, Freyman RL, Keen R.** Manipulations of listeners' echo perception are reflected in event-related potentials. *J Acoust Soc Am* 129: 301–309, 2011.
- Schroeder CE, Smiley J, Fu KG, McGinnis T, O'Connell MN, Hackett TA.** Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int J Psychophysiol* 50: 5–17, 2003.
- Shahin AJ, Kerlin JR, Bhat J, Miller LM.** Neural restoration of degraded audiovisual speech. *Neuroimage* 60: 530–538, 2011.
- Shahin AJ, Roberts LE, Chau W, Trainor LJ, Miller LM.** Music training leads to the development of timbre-specific gamma band activity. *Neuroimage* 41: 113–122, 2008.
- Shahin AJ, Trainor LJ, Roberts LE, Backer KC, Miller LM.** Development of auditory phase-locked activity for music sounds. *J Neurophysiol* 103: 218–229, 2010.
- Smiley JF, Falchier A.** Multisensory connections of monkey auditory cerebral cortex. *Hear Res* 258: 37–46, 2009.
- Spence C, Driver J (Editors).** *Crossmodal Space and Crossmodal Attention*. Oxford, UK: Oxford Univ. Press, 2004.
- Spierer L, Bellmann-Thiran A, Maeder P, Murray MM, Clarke S.** Hemispheric competence for auditory spatial representation. *Brain* 132: 1953–1966, 2009a.
- Spierer L, Bourquin NM, Tardif E, Murray MM, Clarke S.** Right hemispheric dominance for echo suppression. *Neuropsychologia* 47: 465–472, 2009b.
- Stein B, Stanford T, Wallace M, Vaughan JW, Jiang W.** Crossmodal spatial interactions in subcortical and cortical circuits. In: *Crossmodal Space and Crossmodal Attention*, edited by Spence C, Driver J. Oxford, UK: Oxford Univ. Press, 2004, p. 25–50.
- Stein BE, Stanford TR, Rowland BA.** The neural basis of multisensory integration in the midbrain: its organization and maturation. *Hear Res* 258: 4–15, 2009.
- Sumby WH, Pollack I.** Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26: 212–215, 1954.
- Thomas GJ.** Experimental study of the influence of vision on sound localization. *J Exp Psychol* 28: 163–177, 1941.
- Wallach H, Newman EB, Rosenzweig MR.** The precedence effect in sound localization. *Am J Psychol* 62: 315–336, 1949.
- Warner CB, Juola JF, Koshino H.** Voluntary allocation versus automatic capture of visual attention. *Percept Psychophys* 48: 243–251, 1990.
- Whitfield IC, Diamond IT, Chiveralls K, Williamson TG.** Some further observations on the effects of unilateral cortical ablation on sound localization in the cat. *Exp Brain Res* 31: 221–234, 1978.
- Witkin HA, Wapner S, Leventhal T.** Sound localization with conflicting visual and auditory cues. *J Exp Psychol* 43: 58–67, 1952.
- Woods TM, Recanzone GH.** Visually induced plasticity of auditory spatial perception in macaques. *Curr Biol* 14: 1559–1564, 2004.
- Zatorre RJ, Bouffard M, Ahad P, Belin P.** Where is “where” in the human auditory cortex? *Nat Neurosci* 5: 905–909, 2002.

