

Published in final edited form as:

Curr Biol. 2011 February 8; 21(3): 221–225. doi:10.1016/j.cub.2010.12.051.

Visual influences on echo suppression

Christopher W. Bishop^{1,+}, Sam London¹, and Lee M. Miller^{1,2}

¹ University of California, Davis Center for Mind and Brain

² University of California, Davis Department of Neurobiology, Physiology, and Behavior

Summary

Locating sounds in realistic scenes is challenging due to distracting echoes and coarse spatial acoustic estimates. Fortunately, listeners can improve performance through several compensatory mechanisms. First, their brains perceptually suppress short latency (1-10 msec) echoes by constructing a representation of the acoustic environment in a process called the precedence effect [1]. This remarkable ability depends on the spatial and spectral relationship between the first or precedent sound wave and subsequent echoes [2]. In addition to using acoustics alone, the brain also improves sound localization by incorporating spatially precise visual information. Specifically, vision refines auditory spatial receptive fields [3] and can capture auditory perception such that sound is localized towards a coincident visual stimulus [4]. Although visual cues and the precedence effect are each known to improve performance independently, it is not clear if these mechanisms can cooperate or interfere with each other. Here, we demonstrate that echo suppression is enhanced when visual information spatially and temporally coincides with the precedent wave. Conversely, echo suppression is inhibited when vision coincides with the echo. These data show that echo suppression is a fundamentally multisensory process in everyday environments, where vision modulates even this largely automatic auditory mechanism to organize a coherent spatial experience.

Results

In Experiment 1, we quantified the visual contribution to echo suppression by presenting lead-lag auditory pairs – primary-echo pairs of 15 msec noise-bursts – at each subject's echo threshold (5.23 ± 0.55 msec delay). Lead-lag pairs occurred either unimodally (A_{PE}) or accompanied by a 15 msec flash of light at the leading ($A_{PE}V_{Lead}$) or lagging ($A_{PE}V_{Lag}$) location (Figure 1C, left panel). In addition to these Precedence Conditions, several Control Conditions were included to ensure that subjects performed the task and to distinguish any effects from visual capture of audition, known as the ventriloquist's illusion [4]. These conditions included a single auditory noise-burst presented on one side, either unimodally (A) or accompanied by synchronous visual stimulation on the same (AV_{Lead}) or opposite side (AV_{Lag}) (Figure 1C, right panel). Sounds were presented from speakers positioned 18° from midline and approximately 120 cm from the subject's ears. To rule out any idiosyncratic acoustic differences between speakers, we physically switched the speakers half way through the experiment for each subject. Visual stimulation was provided by two light emitting diodes (LEDs) suspended directly above each speaker cone (Figure 1A). For

⁺ corresponding author: Contact: cwbishop@ucdavis.edu, Phone: +1 (530)297-4426, Address: 267 Cousteau Place, Davis, CA 95618.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

each trial, subjects indicated whether they heard a sound from *one-location* or *two-locations* and whether they heard a sound on the *left-side*, *right-side*, or *both-sides*, with two sequential button presses approximately 1.5 sec after the stimuli (Figure 1B). As detailed in the Experimental Procedures, the percentage of *one-location* responses was used to quantify echo suppression. Responses to the hemispace question are summarized in Figure S2A and a full account of response combinations is reported in Table S1.

Vision Influences Echo Suppression

Subjects suppressed the echo, i.e. responded *one-location*, on $55.26 \pm 6.60\%$ of trials in condition A_{PE} , $69.40 \pm 7.57\%$ in $A_{PE}V_{Lead}$, and $45.38 \pm 7.33\%$ in $A_{PE}V_{Lag}$ (Figure 2A). The percentage of *one-location* responses was included as a repeated measure in a three-way within-subject ANOVA with factors [Leading-Side \times Condition \times Speaker-Arrangement].

The main effect of Condition was significant ($p < 0.001$; $\eta_p^2 = 0.47$), while Leading-Side and Speaker Arrangement were not (Leading Side $p = 0.12$, $\eta_p^2 = 0.17$; Left $53.74 \pm 6.30\%$, Right $59.62 \pm 7.35\%$, mean difference (\bar{d}) = $5.88 \pm 3.51\%$; Speaker Arrangement $p = 0.61$; $\eta_p^2 = 0.02$; $57.34 \pm 6.53\%$ versus $56.01 \pm 6.95\%$ when switched, $\bar{d} = 1.33 \pm 2.56\%$). Additionally, Speaker-Arrangement did not interact with any other factor ($p > 0.44$), suggesting that the speakers, with their virtually indistinguishable response functions (see Experimental Procedures), were behaviorally equivalent. Post-hoc tests revealed a significant difference between all conditions (Figure 2B; $A_{PE}V_{Lead} > A_{PE} > A_{PE}V_{Lag}$; pair-wise comparisons with Fisher's LSD). In other words, echo suppression increased by $14.15 \pm 3.28\%$ ($p < 0.006$; $\eta_p^2 = 0.57$) with synchronous visual stimulation on the leading side ($A_{PE}V_{Lead}$) and decreased by $9.88 \pm 5.08\%$ ($p = 0.049$; $\eta_p^2 = 0.21$) on the lagging side ($A_{PE}V_{Lag}$) compared to a unimodal lead-lag pair (A_{PE}).

Subjects performed well in all Control Conditions, exceeding 94% accuracy (Figure 2A). The percentage of *one-location* responses in the three Control Conditions (A, AV_{Lead} , AV_{Lag}) was included in a separate three-way, within-subject ANOVA with factors [Leading-Side \times Condition \times Speaker-Arrangement]. This analysis revealed that subjects performed better with sounds presented on the Right-Side ($97.48 \pm 0.90\%$) than the Left-Side ($94.86 \pm 1.00\%$) ($p = 0.01$; $\eta_p^2 = 0.38$; $\bar{d} = 2.62 \pm 0.89\%$). However, there were no other significant main effects or interactions among control conditions ($p > 0.05$). Importantly, subjects rarely mislocalized sounds when visual stimulation was presented in the opposite hemifield (AV_{Lag} , Figure S2). Under different circumstances, namely if the spatial disparity between sight and sound were much less than our 36° [5], vision might capture audition in the ventriloquist's illusion [4]. This key control therefore indicates that our observed effects are not due to visual spatial capture but rather reflect genuine visual modulation of the echo's perceptual salience.

Visual Influence is Robust to Short-term Learning

Previous work has suggested that audiovisual interactions [6] and echo suppression [7] may be altered with extended practice. Similarly, we thought it possible that vision's contribution to echo suppression may change over time. The within-subject design of Experiment 1 allowed us to test this hypothesis by dividing each subject's data in half and including the percentage of *one-location* responses in a three factor repeated measure ANOVA with [Leading-Side \times Condition \times Session (First/Second)] as within-subject factors. The analysis revealed no main effect of Session ($p = 0.92$; $\eta_p^2 = 0.0008$; First $56.54 \pm 6.10\%$, Second $56.81 \pm 7.33\%$, $\bar{d} = 0.27 \pm 2.58\%$) or interactions between Session and other factors ($p > 0.34$); this suggests that echo suppression did not change with extended practice. More importantly, a

lack of a Session \times Condition interaction suggests that vision's contribution to echo suppression is a stable phenomenon and resistant to short-term learning effects.

Visually Induced Suppression Depends on Audiovisual Temporal Alignment

Although Experiment 1 demonstrates that visual stimulation is sufficient to affect echo suppression, it is not clear from these data whether temporal coincidence is strictly necessary. Therefore in Experiment 2, we repeated Experiment 1 with 18 independent subjects and added stimuli in the $A_{PE}V_{Lead}$ and $A_{PE}V_{Lag}$ configurations with the visual stimulus delayed by 100 msec ($A_{PE}V_{Lead/Lag(+100)}$) or 400 msec ($A_{PE}V_{Lead/Lag(+400)}$). Importantly, due to well documented effects of crossmodal, exogenous attentional cues on auditory processing [8], we deliberately excluded visual-leading stimulus configurations to avoid ambiguous interpretations. Based on existing studies suggesting a narrow (~ 200 msec) window of temporal integration for short duration audiovisual stimuli (e.g. [9]), we hypothesized that vision's influence over echo suppression would be eliminated with a 400 msec offset.

Echo thresholds in Experiment 2 were estimated to be 4.31 ± 0.46 msec and resulted in $50.81 \pm 4.18\%$ *one-location* responses in condition A_{PE} , $70.60 \pm 3.27\%$ in condition $A_{PE}V_{Lead}$, and $48.03 \pm 4.47\%$ for $A_{PE}V_{Lag}$ (Figure 3A). The reader will recall that these conditions are identical to those in Experiment 1. The percentage of *one-location* responses was included in a two-factor [Leading-Side \times Condition] repeated measure ANOVA. The main effect of Condition was significant (Figure 3B; $p < 0.001$; $\eta_p^2 = 0.63$) while Leading-Side was not ($p = 0.10$; $\eta_p^2 = 0.15$; Left $50.23 \pm 5.71\%$, Right $62.73 \pm 4.23\%$, $d = 12.50 \pm 7.14\%$), nor was the interaction significant ($p = 0.98$; $\eta_p^2 = 0.001$). As in Experiment 1, post-hoc tests revealed greater echo suppression in condition $A_{PE}V_{Lead}$ compared to A_{PE} (Figure 3B; $p < 0.001$; $\eta^2 = 0.71$; mean increase of $19.79 \pm 3.05\%$). However, in contrast to Experiment 1, there was no significant difference between conditions $A_{PE}V_{Lag}$ and A_{PE} (Figure 3B; $p = 0.42$; $\eta^2 = 0.04$; $d = 2.78 \pm 3.38\%$). In other words, visual stimulation at the echo location failed to inhibit echo suppression as it did in Experiment 1. This null result could be due to the context of the task (e.g. the addition of temporally offset stimuli) or an unrepresentative sampling of the population. We took several steps to address this apparent discrepancy. First, we included the percentage of *one-location* responses for conditions A_{PE} , $A_{PE}V_{Lead}$, and $A_{PE}V_{Lag}$ from all 33 subjects in Experiments 1 and 2 into a single, two-factor [Leading-Side \times Condition] repeated measure ANOVA. The main effect of Condition was significant ($p < 0.001$; $\eta_p^2 = 0.55$; A_{PE} $52.63 \pm 3.80\%$, $A_{PE}V_{Lead}$ $70.03 \pm 3.87\%$, and $A_{PE}V_{Lag}$ $46.14 \pm 4.27\%$); post-hoc tests revealed a significant difference between all conditions ($A_{PE}V_{Lead} > A_{PE} > A_{PE}V_{Lag}$), with a $17.4 \pm 2.24\%$ increase ($p < 0.001$; $\eta^2 = 0.65$) in condition $A_{PE}V_{Lead}$ and a $6.49 \pm 2.96\%$ decrease ($p = 0.02$; $\eta^2 = 0.13$) in condition $A_{PE}V_{Lag}$ compared to A_{PE} . Additionally, we replicated Experiment 1 in an independent set of subjects ($N = 12$) and observed virtually identical results (Figure S3). Together, these experiments provide evidence across 45 independent subjects that vision's effect on echo suppression is a robust and replicable phenomenon.

Finally, to assess the temporal dependence of vision's contribution to echo suppression, we included the percentage of *one-location* responses for conditions A_{PE} , $A_{PE}V_{Lead}$, $A_{PE}V_{Lead(+100)}$, and $A_{PE}V_{Lead(+400)}$ in a two-factor [Leading-Side \times Condition] repeated measure ANOVA. Due to a weak, insignificant effect in condition $A_{PE}V_{Lag}$, we did not have a dynamic range over which to test for a temporal dependence of visual stimulation on the lag side. As a result, we focused on the consequences of visual temporal offsets on the

leading side. The main effect of Condition (Figure 3) was significant ($p < 0.001$; $\eta_p^2 = 0.73$; $A_{PE} V_{Lead} 50.81 \pm 4.17\%$, $A_{PE} V_{Lead} 70.60 \pm 3.27\%$, $A_{PE} V_{Lead (+100)} 73.95 \pm 3.25\%$, and

$A_{PE} V_{Lead (+400)} 54.75 \pm 4.67\%$) while Leading-Side was not ($p = 0.09$; $\eta_p^2 = 0.16$; Left $56.37 \pm 5.48\%$, Right $68.69 \pm 4.36\%$; $\bar{d} = 12.32 \pm 6.81\%$). Post-hoc tests of Condition revealed that visually enhanced echo suppression was virtually eliminated by a 400 msec temporal offset, with $A_{PE} V_{Lead}, A_{PE} V_{Lead (+100)} > A_{PE}, A_{PE} V_{Lead (+400)}$. Compared to A_{PE} , condition $A_{PE} V_{Lead}$ increased by $19.79 \pm 3.05\%$ ($p < 0.001$; $\eta^2 = 0.71$), $A_{PE} V_{Lead (+100)}$ by $23.15 \pm 1.67\%$ ($p < 0.001$; $\eta^2 = 0.92$), and $A_{PE} V_{Lead (+400)}$ by $3.94 \pm 2.38\%$ ($p = 0.11$; $\eta^2 = 0.14$) (Figure 3B). That is, a visually induced increase in echo suppression critically depends on gross audiovisual temporal alignment.

Discussion

The brain employs many strategies to improve spatial perception in reverberant conditions, including suppressing echoes and exploiting complementary visual spatial information. To date, the contributions of these two mechanisms have only been considered in isolation and thus the role of visual information on echo suppression was unknown. Our results over several experiments clearly demonstrate that vision can affect echo suppression, likely by providing crossmodal evidence of an object's existence and location in space. Furthermore, this interaction is robust to short-term learning effects and critically depends on audiovisual temporal alignment, suggestive of an early audiovisual interaction. In the following sections, we discuss the potential origins, implications, and neural basis for this phenomenon as well as how these findings impact our understanding of multisensory spatial processing generally.

This crossmodal interaction raises several possible explanations. The first and most likely is that audiovisual inputs are integrated to enhance the neural representation of the leading or lagging sound thereby increasing its perceptual salience [10]. This integration hypothesis is supported by the temporal dependence of the effect. We demonstrate that vision's contribution to echo suppression is virtually eliminated with a 400 msec offset, consistent with previous reports suggesting a short (~ 200 msec) window for temporal integration (e.g. [9] but see [11]). An alternative explanation is that the modulation of echo suppression results from visually-cued, rapid deployment of spatial attention that then affects target detectability. However attentional effects are unlikely, due to the relatively slow nature of attentional deployment and the tight temporal constraints of the stimuli. For instance, attention cannot explain a robust reduction in echo suppression with temporally coincident visual stimulation at the echo location since there was no time for subjects to deploy attentional resources in response to the exogenous visual cue. As a result, we argue that this phenomenon is due to early (low-level) integration of audiovisual information.

Vision's contribution to echo suppression may have practical behavioral consequences for both sound localization and general acoustic processing in reverberant environments. For instance, studies have demonstrated that subjects understand speech better when echoes are suppressed by “building-up” a representation of acoustic space [12]. Visually enhanced echo suppression may have similar behavioral consequences that, to our knowledge, have not been tested. Second, a reduction in echo suppression with visual stimulation at the same time and place as the echo may serve as a putative override signal to relatively automatic auditory echo suppression. Practically speaking, such a mechanism would prevent erroneous suppression of temporally proximal sounds originating from independent objects in space and instead provide crossmodal evidence of independent auditory sources that are subsequently organized into auditory streams [13].

Finally, our findings help formulate testable hypotheses about the neural basis of echo suppression in everyday reverberant environments. First, a visual contribution to the precedence effect suggests the involvement of a previously unconsidered multisensory neural substrate(s). Candidates include classic subcortical and cortical multisensory regions, such as the superior colliculus [14] and superior temporal sulcus [15]; alternatively, vision may exert its influence in what is traditionally considered unimodal auditory cortex, as recent studies have demonstrated anatomical connections between early auditory and visual areas [16,17] as well as early functional consequences [18]. Additionally, a persistent interaction with a 100 msec temporal offset suggests that echo suppression is a slow, progressive process. This notion contradicts existing models claiming that echo suppression is a necessarily early or “automatic” auditory-only process likely occurring in the inferior colliculus (IC) (see [2] for review). Instead, our data support electroencephalography (EEG) studies in humans indicating that echo acoustics are still represented veridically in the auditory brainstem [19,20]. In these studies, neural correlates of echo suppression first manifest around the time sensory responses reach cortex (~30 msec) and these continue for two hundred milliseconds (e.g. auditory potentials P1, N1, and P2 [21,22]). This allows ample time for extensive, direct neural influence on echo suppression via visual and multisensory cortices.

In conclusion, our data show that vision affects even a traditionally unimodal, exquisitely time-sensitive auditory mechanism for parsing spatial scenes. This supports the general view that spatial hearing in realistic environments is fundamentally multisensory. Future investigations into this phenomenon might address its precise temporal and spatial dependencies, its behavioral consequences, and the neural substrates through which it is realized.

Experimental Procedures

Subjects

In accordance with procedures approved by the University of California, a total of 45 subjects (Experiment 1: 20 subjects, 13 female, mean age of 22 yrs., range 19-30 yrs; Experiment 2: 25 subjects, 13 female, mean age of 20 yrs, range 18-25 yrs) gave written consent prior to their participation. Subjects had self-reported good hearing and normal or corrected-to-normal vision. Seven subjects (two in Experiment 1 and five in Experiment 2) were excluded prior to data collection because their echo thresholds were below 2 msec. A 2 msec echo threshold cutoff was imposed to ensure that subjects clearly lateralized suppressed sounds to the location of the leading speaker. This cutoff was determined through testing of three individuals (see Figure S1); these findings generally agree with other reports (e.g. [2]). An additional three subjects were excluded from data analysis in Experiment 1 because their performance in one or more control conditions fell below a predefined cutoff of 70% *one-location* responses, indicating that they were not performing the task. As a result, the data reported are based on 15 (10 female) and 18 (9 female) subjects for Experiment 1 and 2, respectively.

Stimuli and Task

Subjects sat approximately 120 cm from a computer monitor (Dell FPW2407) in an acoustically transparent chair (Herman Miller AE500P; www.hermanmiller.com) in a double-walled, acoustically dampened chamber. Sounds were presented from a set of Tannoy Precision 6 (www.tannoy.com) speakers positioned 18° to the left and right of midline. Visual stimulation was provided by two white light emitting diodes (LEDs, 7 candelas) positioned at the top edge of the speaker cone (see Figure 1A). To enhance the perceptual salience and visual angle of the LEDs, each LED was placed in a white ping-

pong ball to produce diffuse light flashes over an approximately 1.7° visual angle. Auditory stimuli were 15 msec noise bursts with 0.5 msec linear onset and offset ramps. The linear ramps were set to 0.5 msec to approximately match the rise time of the LEDs measured using a photodiode (THORLABS DET36A, www.thorlabs.us) and oscilloscope (Tektronix TDS 2004; www2.tek.com). Visual stimuli were 15 msec light flashes generated using the LEDs. Auditory stimuli were presented at approximately 75 dB(A) from each speaker and calibrated prior to each session using a sound pressure level (SPL) meter. Temporally leading and lagging auditory pairs were presented from the speakers to simulate a primary wave and its corresponding echo. Unique noise bursts were generated for each trial but lead-lag pairs consisted of identical noise samples for a given trial.

In Experiment 1, precedence effect stimuli (lead-lag noise bursts) were either presented in the absence of a visual stimulus (A_{PE}) or accompanied by a visual flash spatiotemporally aligned with the leading ($A_{PE}V_{Lead}$) or lagging ($A_{PE}V_{Lag}$) sound (Figure 1C, left panel). Additional control conditions were included to ensure that subjects performed the task and that their responses were not significantly biased by the visual stimulus. These included an auditory stimulus presented from a single speaker unimodally (A) or accompanied by a flash of light at the same (AV_{Lead}) or the opposite side (AV_{Lag}) as the auditory stimulus (Figure 1C, right panel). Experiment 2 consisted of these conditions as well as stimuli presented in the $A_{PE}V_{Lead/Lag}$ configuration with the visual stimulus delayed by 100 msec ($A_{PE}V_{Lead/Lag(+100)}$) and 400 msec ($A_{PE}V_{Lead/Lag(+400)}$) relative to sound onset.

Each trial began with stimulus delivery (<25 msec total) followed by a delay (1.5 sec) prior to subject responses. During earlier pilot work, we found that subject responses often did not accurately represent their perception in a speeded response paradigm. The forced delay was introduced to counteract these effects and ensure response accuracy. Stimulus onset asynchronies were selected randomly from a uniform distribution ranging from 2.5 to 2.9 sec. Subjects were asked to answer two questions for each trial. First, they indicated which side they heard a sound on by pressing their right index (*left-side*), middle (*both-sides*), or ring (*right-side*) fingers. Second, they indicated how many different locations they heard a sound from by pressing their left index (*one-location*) and middle (*two-location*) fingers. Although these two questions ultimately provided nearly identical information in the current study (see Results), both questions were included to allow the subject to classify “intermediate percepts”: when all acoustic energy is perceived in one hemifield of space (e.g. *left-side*) but at two spatially distinct locations (i.e. *two-locations*) [2,21]. Without both response options, subjects would not have been able to accurately describe their perception, thereby confounding our interpretations. All stimulus presentation and response recording was coordinated through Neurobehavioral System's Presentation Software (www.neuro-bs.com). Stimuli were dynamically generated prior to the start of each trial in MATLAB (www.mathworks.com).

Experiment 1 consisted of three distinct parts: training, calibration, and the experimental task. During training, subjects were presented with a verbally narrated PowerPoint presentation to familiarize the subject with the stimuli and task. Importantly, subjects were told that their task only involved what they heard and were encouraged to ignore the flashing lights throughout training. The PowerPoint presentation was followed by several examples of each stimulus type: A, A_{PE} , $A_{PE}V_{Lead}$, $A_{PE}V_{Lag}$, AV_{Lead} , and AV_{Lag} . Leading and lagging pairs were presented with temporal lags that were clearly suppressed (2 msec) or clearly not suppressed (20 and 100 msec). Subjects then performed a brief (3 min) mock run that included stimuli in all conditions with equal frequency. The temporal offset for lead-lag pairs was set to 100 msec during the mock run to provide clear examples of likely response categories. Training for Experiment 2 was virtually identical, but had additional examples of conditions $A_{PE}V_{Lead/Lag(+100)}$ and $A_{PE}V_{Lead/Lag(+400)}$.

A calibration session immediately followed training. Stimuli were presented in the A_{PE} condition to identify quickly each subject's echo threshold, defined here as the temporal lag at which a subject suppresses the echo (i.e. responds *one-location*) on 50% of trials, using a one-up-one-down staircase algorithm. The algorithm was initiated with a 2 msec lag and incremented or decremented the temporal offset of lead-lag pairs by 0.5 msec for suppressed (*one-location*) and not-suppressed (*two-locations*) stimuli, respectively. The algorithm terminated after the direction of change reversed 10 times. Subjects were presented with approximately the same number of left- and right-leading sounds pseudorandomly such that every 6 trials consisted of 3 left- and right-leading sounds. After the calibration session, the experimenter played a series of test stimuli for the subject and solicited verbal reports from the subject to ensure accurate echo threshold estimation. In Experiment 2, we implemented a short "threshold check" where subjects performed the experimental task at their estimated echo threshold for eight trials each of conditions A, A_{PE} , $A_{PE}V_{Lead}$, $A_{PE}V_{Lag}$, AV_{Lead} , and AV_{Lag} . The subject's echo threshold was then adjusted as necessary to better target 50% echo suppression in condition A_{PE} . Importantly, conditions $A_{PE}V_{Lead/Lag(+100)}$ and $A_{PE}V_{Lead/Lag(+400)}$ were excluded from this procedure to prevent experimenter bias.

Subjects participated in the main experiment following the calibration session. In Experiment 1, stimuli were presented in all conditions for a total of 100 trials per condition (50 for Left- and Right-Leading). The task was identical to the training and calibration sessions, consisted of 600 trials, and lasted approximately 30 minutes. Subjects were given an optional break every 6-8 minutes. To rule out potential differences between speakers, the experimenter physically switched the speakers after the first 15 min (300 trials) of testing and included Speaker-Arrangement as a within-subject factor (see Results). In Experiment 2, subjects performed the task for three, approximately 7 min sessions consisting of eight trials per condition and side for a total of 24 trials per condition per side.

Speaker Response Functions

The impulse responses for the Tannoy Precision 6 speakers used in these experiments were acquired using a SHURE KSM44 omnidirectional microphone (www.shure.com). A single-sample impulse was played through each speaker and recorded at 96-kHz 10 times and temporally averaged. The impulse response revealed a single prominent echo arriving approximately 5.8 msec later and 9.10 dB quieter than the primary wave. Frequency response functions were acquired by playing a single 15 msec white-noise burst (see above) 10 times from each speaker. These recordings were temporally averaged and their spectral density estimated using Welch's method in MATLAB. The spectral densities revealed that the two speakers were virtually indistinguishable (<3 dB difference) up to 20 kHz.

Statistical Analysis and Reporting

For all reported repeated-measure ANOVAs, we chose the percentage of *one-location* responses as the dependent measure. This response category was used because it most accurately indexes complete echo suppression. Additionally, subjects tended not to respond one-side (e.g. *left-side*), *two-locations*, thus the responses indicative of suppression (*one-location* or leading-side) in either question were highly correlated. All statistical quantification was done in STATISTICA version 8.0 (www.statsoft.com). Reported p-values are corrected for non-sphericity using Greenhouse-Geisser correction where appropriate and all post-hoc, pair-wise comparisons were performed using Fisher's Least Significant Difference (LSD). Statistical significance was assessed with $\alpha=0.05$. With extremely small p-values, we specified the range to within three decimal places. Eta-squared (η^2) or Partial eta-squared (η_p^2) is reported as a measure of effect size. Mean differences are denoted as \bar{d} [23]. Unless noted otherwise, means and mean differences are reported with standard error of the mean (SEM).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Jess Kerlin for his technical assistance and four referees for their insightful comments and rigorous standards. This research was supported by R01-DC008171 and 3R01-DC008171-04S1 (LMM) and F31-DC011429 (CWB) awarded by the National Institute on Deafness and other Communication Disorders (NIDCD). 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.

References

1. Wallach H, Newman EB, Rosenzweig MR. The precedence effect in sound localization. *The American journal of psychology*. 1949; 62:315–336. [PubMed: 18134356]
2. Litovsky RY, Colburn HS, Yost WA, Guzman SJ. The precedence effect. *The Journal of the Acoustical Society of America*. 1999; 106:1633–1654. [PubMed: 10530009]
3. Bizley JK, King AJ. Visual-auditory spatial processing in auditory cortical neurons. *Brain Res*. 2008; 1242:24–36. [PubMed: 18407249]
4. Howard, IP.; Templeton, WB. *Human Spatial Orientation*. New York: Wiley; 1966.
5. Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*. 2001; 12:7–10. [PubMed: 11201094]
6. Powers AR 3rd, Hillock AR, Wallace MT. Perceptual training narrows the temporal window of multisensory binding. *J Neurosci*. 2009; 29:12265–12274. [PubMed: 19793985]
7. Saberi K, Perrott DR. Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. *The Journal of the Acoustical Society of America*. 1990; 87:1732–1737. [PubMed: 2341678]
8. Spence, C.; Driver, J., editors. *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press; 2004.
9. Zampini M, Guest S, Shore DI, Spence C. Audio-visual simultaneity judgments. *Perception & psychophysics*. 2005; 67:531–544. [PubMed: 16119399]
10. Stein, B.; Stanford, T.; Wallace, M.; Vaughan, JW.; Jiang, W. Crossmodal spatial interactions in subcortical and cortical circuits. In: Spence, C.; Driver, J., editors. *Crossmodal space and crossmodal attention*. New York: Oxford; 2004. p. 25-50.
11. Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the crossmodal integration of speech. *J Neurosci*. 2005; 25:5884–5893. [PubMed: 15976077]
12. Brandewie E, Zahorik P. Prior listening in rooms improves speech intelligibility. *The Journal of the Acoustical Society of America*. 2010; 128:291–299. [PubMed: 20649224]
13. Bregman, AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press; 1994.
14. Meredith MA, Stein BE. Interactions among converging sensory inputs in the superior colliculus. *Science (New York, NY)*. 1983; 221:389–391.
15. Calvert GA. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex*. 2001; 11:1110–1123. [PubMed: 11709482]
16. Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF. Projection from Visual Areas V2 and Prostriata to Caudal Auditory Cortex in the Monkey. *Cereb Cortex*. 2009
17. Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci*. 2002; 22:5749–5759. [PubMed: 12097528]
18. Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain research*. 2002; 14:187–198. [PubMed: 12063142]
19. Liebenthal E, Pratt H. Human auditory cortex electrophysiological correlates of the precedence effect: Binaural echo lateralization suppression. *The Journal of the Acoustical Society of America*. 1999; 106:291–303.

20. Damaschke J, Riedel H, Kollmeier B. Neural correlates of the precedence effect in auditory evoked potentials. *Hear Res.* 2005; 205:157–171. [PubMed: 15953525]
21. Backer KC, Hill KT, Shahin AJ, Miller LM. Neural time course of echo suppression in humans. *J Neurosci.* 2010; 30:1905–1913. [PubMed: 20130199]
22. Sanders LD, Joh AS, Keen RE, Freyman RL. One sound or two? Object-related negativity indexes echo perception. *Perception & psychophysics.* 2008; 70:1558–1570. [PubMed: 19064498]
23. Zar, JH. *Biostatistical Analysis*. 3rd. New Jersey: Prentice Hall; 1996.

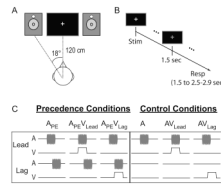


Figure 1.

Apparatus and stimuli. (A) Auditory stimuli were presented from two speakers positioned 18° to the right and left of midline. Visual stimuli were presented via light emitting diodes (LEDs, white circles) suspended just above the speaker cone. Auditory and visual stimuli were 15 msec in duration. (B) Trial structure. Subjects waited 1.5 sec after stimulus presentation to respond. The beginning of the response window was marked by a change from a white to a green fixation cross (shown here as gray). Subjects were given 1.0 – 1.4 sec to respond before the next stimulus presentation. (C) Stimulus conditions. All precedence effect conditions and control conditions are graphically depicted. Temporally leading and lagging auditory pairs were presented alone (A_{PE}) or with a visual stimulus spatially and temporally aligned with the leading ($A_{PE}V_{Lead}$) or lagging ($A_{PE}V_{Lag}$) stimulus. Control conditions included a single noise burst presented alone (A) or with visual stimulation on the same (AV_{Lead}) or opposite side (AV_{Lag}). Stimulus (Stim), Response (Resp), Auditory (A), Visual (V).

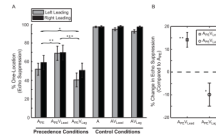


Figure 2.

Vision contributes to auditory echo suppression. (A) Mean percentage of *one-location* responses (echo suppression) is plotted as a function of condition and side. (B) Echo suppression was enhanced in condition $A_{PE}V_{Lead}$ (squares) and inhibited in condition $A_{PE}V_{Lag}$ (circles) compared to condition A_{PE} . Naming conventions are identical to Figure 1. Brackets and asterisks indicate significant differences between Precedence Conditions (* $p=0.049$, ** $p=0.006$, *** $p<0.001$). (N=15; mean \pm SEM)

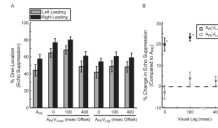


Figure 3.

Visually enhanced echo suppression critically depends on gross temporal alignment. (A) The mean percentage of *one-location* responses (echo suppression) is plotted as a function of condition and side. Naming conventions are identical to those in Figure 1. Integer values along the abscissa represent the delay (msec) of the visual stimulus relative to the onset of the spatially coincident auditory stimulus. (B) The difference in echo suppression between conditions $A_{PE}V_{Lag}$ and A_{PE} relative to A_{PE} at 0, 100, and 400 msec visual lag. Importantly, the visual contribution to echo suppression is eliminated with a 400 msec offset. Symbols are offset at 400 msec delay for clarity. (* $p < 0.001$) (N=18; mean \pm SEM)