



Published in final edited form as:

*Exp Brain Res.* 2017 February ; 235(2): 585–595. doi:10.1007/s00221-016-4820-4.

## Accumulation and Decay of Visual Capture and the Ventriloquism Aftereffect Caused by Brief Audio-Visual Disparities

Adam K. Bosen<sup>1</sup>, Justin T. Fleming<sup>2</sup>, Paul D. Allen<sup>2</sup>, William E. O'Neill<sup>2</sup>, and Gary D. Paige<sup>2</sup>

<sup>1</sup>Department of Biomedical Engineering, University of Rochester, Rochester, NY

<sup>2</sup>Department of Neuroscience, University of Rochester, Rochester, NY

### Abstract

Visual capture and the ventriloquism aftereffect resolve spatial disparities of incongruent auditory-visual (AV) objects by shifting auditory spatial perception to align with vision. Here, we demonstrated the distinct temporal characteristics of visual capture and the ventriloquism aftereffect in response to brief AV disparities. In a set of experiments, subjects localized either the auditory component of AV targets (*A within AV*) or a second sound presented at varying delays (1-20s) after AV exposure (*A2 after AV*). AV targets were trains of brief presentations (1 or 20), covering a  $\pm 30^\circ$  azimuthal range, and with  $\pm 8^\circ$  (R or L) disparity. We found that the magnitude of visual capture generally reached its peak within a single AV pair and did not dissipate with time, while the ventriloquism aftereffect accumulated with repetitions of AV pairs and dissipated with time. Additionally, the magnitude of the auditory shift induced by each phenomenon was uncorrelated across listeners and visual capture was unaffected by subsequent auditory targets, indicating that visual capture and the ventriloquism aftereffect are separate mechanisms with distinct effects on auditory spatial perception. Our results indicate that visual capture is a 'sample-and-hold' process that binds related objects and stores the combined percept in memory, whereas the ventriloquism aftereffect is a 'leaky integrator' process that accumulates with experience and decays with time to compensate for cross-modal disparities.

### Keywords

visual capture; ventriloquism; ventriloquism aftereffect; auditory; visual; cross-modal

### Introduction

Vision and hearing provide complementary information about the world around us. Under natural conditions, objects in the environment often produce both visual images and related sounds. It is generally advantageous to integrate these auditory and visual (AV) cues, as it provides more information about the object than either sense provides individually, and allows for perceptual corrections if the two senses are not in alignment with one another. Spatially misaligned auditory and visual cues elicit two perceptual phenomena known as

visual capture and the ventriloquism aftereffect (Chen and Vroomen, 2013). Visual capture (also known as the “ventriloquism effect”, Howard and Templeton, 1966) occurs when the sensed locations of related auditory and visual targets originate from spatially disparate locations. For individuals with normal sensation, vision is typically much more spatially precise than audition. As a result, it tends to dominate the perceived common location, and thereby “captures” the auditory percept, which results in a strong bias in perceived auditory location toward the visual target (Jack and Thurlow, 1973; Thurlow and Jack, 1973). In comparison, the ventriloquism aftereffect occurs when an auditory target is presented in isolation following spatially disparate pairs of auditory and visual targets. In this case, the sensory disparity between target locations in the audio-visual pair is used to adjust the perceived location of the subsequent isolated auditory target (Radeau and Bertelson, 1974). The purpose of this paper is to contrast the dynamic effects of visual capture and the ventriloquism aftereffect on auditory spatial perception in response to brief trains of spatially disparate auditory and visual targets.

Visual capture occurs because sensory perception is inherently uncertain, and vision is typically more precise than audition for spatial perception. When there is a perceived disparity between auditory and visual target locations, it is possible that the perceived disparity is caused by sensory error, in which case the targets should be bound into a coherent percept. Alternatively, it is possible that the perceived disparity is caused by an actual physical disparity between the targets, in which case they should be segregated. If the targets are bound together, then the auditory and visual signals are perceived to originate from the average of the auditory and visual locations, weighted by the relative reliability of each sense (Battaglia et al., 2003). If the targets are segregated, then the auditory and visual target locations can be independently reported with little bias in perceived auditory location (Hairston et al., 2003; Wallace et al., 2004). This process of binding or segregating cues across sensory modalities is generally referred to as causal inference (for review, see Shams and Beierholm, 2010) and is not specific to audio-visual interactions. Generally, visual capture occurs with a single exposure to an audio-visual target pair (Hairston et al., 2003; Wallace et al., 2004), so it does not require repetition or familiarity with the targets to occur. If the spatial disparity between targets is relatively small (generally less than 10-15 degrees in azimuth, depending on the study), subjects are often unaware the disparity exists (Bertelson and Radeau, 1981; Radeau and Bertelson, 1974) and visual capture is likely to occur.

In contrast, the ventriloquism aftereffect adjusts auditory spatial perception *after* spatial misalignment between auditory and visual targets. When subjects are exposed to audio-visual spatial disparities, the ventriloquism aftereffect responds by shifting the perceived location of subsequent isolated auditory targets toward the visual location in the spatially disparate audio-visual target pair. The term “ventriloquism aftereffect” is a misnomer; although visual capture (i.e. “ventriloquism”) increases the magnitude of the shift in perceived location of subsequent auditory targets, some shift in perceived location still occurs even when visual capture does not. Examples of this include shifts in auditory perception when the preceding audio-visual pairs are not perceived as originating from the same location (Wozny and Shams, 2011), when subjects are explicitly shown that targets do not arise from common locations (Radeau and Bertelson, 1974; Radeau and Bertelson,

1978), and when subjects are directing their attention to a competing visual task unrelated to the audio-visual target pairs (Eramudugolla et al., 2011).

The ventriloquism aftereffect is not the only mechanism that can alter auditory spatial perception. Maintaining gaze that is eccentric relative to the head in a dark room also produces uniform shifts in auditory space in the direction of fixation (Cui and Razavi, 2010; Dobрева et al., 2012; Razavi et al., 2007), which is referred to as the oculomotor effect. The oculomotor effect is believed to be caused by passive adaptation of the perceived eye-in-head orientation toward straight-ahead over time, which causes visually guided localization of head-referenced targets (such as auditory targets) to be misaligned (Razavi et al., 2007). The ventriloquism aftereffect occurs across a mixture of eye-centered and head-centered reference frames (Kopco et al., 2009), so any misalignment of the eye-to-head coordinate transform, as caused by the oculomotor effect, would confound estimates of the ventriloquism aftereffect. Specifically, if the distribution of visual target locations is biased toward one side and subjects are allowed to look at the visual targets, then both the ventriloquism aftereffect and the oculomotor effect would have similar effects on auditory localization that could not be distinguished from one another. This confounding issue likely affects the data obtained by some experiments intended to measure the ventriloquism aftereffect (Frissen et al., 2003; Frissen et al., 2005; Frissen et al., 2012; Mendonça et al., 2015). In particular, the time course and magnitude of auditory shifts in experiments intended to measure the ventriloquism aftereffect (Frissen et al., 2012) is very similar to the time course and magnitude of auditory shifts observed in the oculomotor effect (Razavi et al., 2007). These similarities across studies suggest that the time course of the ventriloquism aftereffect without the confounding influence of the oculomotor effect is relatively unexplored.

Recent work that did control for eye position suggests that the ventriloquism aftereffect may comprise two distinct mechanisms operating at separate time scales (Bruns and Röder, 2015). Specifically, extensive exposure to fixed audio-visual disparities over hundreds or thousands of repetitions of the disparity produces an enduring shift in auditory spatial perception (Recanzone, 1998; Lewald, 2002; Wozny and Shams, 2011a), indicating that auditory recalibration occurred. In contrast, individual spatially disparate audio-visual targets 35 ms long are sufficient to produce small but significant shifts in auditory spatial perception (Wozny and Shams, 2011). If we extrapolate from the short-term auditory shift data in Wozny and Shams (2011), we would expect the ventriloquism aftereffect to completely compensate for a fixed audio-visual disparity in under 100 exposures to a disparity. However, experiments with extensive exposure to a fixed audio-visual disparity often show only partial recalibration, either in artificial conditions with hundreds or thousands of controlled exposures to a fixed disparity (Lewald, 2002; Wozny and Shams, 2011a), or in natural environments while wearing visual prisms for hours or days (Zwiers et al., 2003; Cui et al., 2008). This indicates that the auditory shifts observed after extensive exposure to a fixed audio-visual disparity may not reflect the same mechanism that shifts auditory perception after a brief audio-visual disparity, although both are commonly referred to in the literature as recalibration or the ventriloquism aftereffect.

The purpose of this work is to measure the accumulation and decay of the ventriloquism aftereffect for brief (1 or 20 repetitions) exposure to audio-visual disparities, while controlling for the oculomotor effect. Wozny and Shams (2011) demonstrated that the auditory shift induced by a single exposure to an audio-visual disparity is attenuated by subsequent audio-visual exposures, but did not determine whether this attenuation is caused by the time delay after the initial disparity or the subsequent exposures themselves. Additionally, repeated exposures to disparities with the same relative position of auditory and visual targets (e.g. visual target always left or always right of auditory) appeared to increase the magnitude of the effect, but the magnitude of the disparity and target location were not fixed across exposures. To expand on this previous work and directly address the time course of the ventriloquism aftereffect, we conducted a series of experiments to demonstrate the accumulation and decay of the ventriloquism aftereffect caused by brief, spatially disparate trains of auditory and visual targets, while controlling for eye position during target presentation to avoid the confounding oculomotor effect. The ventriloquism aftereffect is compared to visual capture under matched conditions within the same subjects, highlighting the different effects each mechanism has on auditory spatial perception. By comparing visual capture and the ventriloquism aftereffect under matched conditions, we can identify the unique temporal properties of each phenomenon and thereby differentiate the effects of cross-modal integration and adaptation on audio-visual spatial perception, elucidating the rapid temporal dynamics of the perceptual mechanisms that adjust cross-modal sensory congruence.

## Materials and Methods

### Subjects

Eighteen volunteers (6 male, ages 18-27 years) recruited from the Rochester community participated in these experiments. All subjects were screened for clinically normal hearing (thresholds less than 20 dB HL, at octave frequencies from 250 Hz - 8 kHz) and normal (or corrected to normal) vision. Ten subjects completed experiment 1, 10 completed experiment 2, and 6 completed experiment 3, with partial overlap between them.

Experimental protocols were approved by the Institutional Review Board at the University of Rochester and were performed in accordance with the 1964 Declaration of Helsinki. All subjects gave informed consent and were compensated for their participation.

### Apparatus

Experiments were conducted in a dark, sound-attenuated chamber designed for the presentation of auditory and visual targets from a range of locations in space (Figure 1a, for additional detail, see Zwiers et al., 2003). Subjects were seated 2 meters from 2 speakers, each mounted on a mobile robotic arm, hidden from sight by acoustically transparent speaker cloth.

Subjects were head-restrained using a custom bite-bar, which was oriented for each subject such that Reid's baseline was earth-horizontal and the subject's cyclopean eye (midway between the eyes) was pointed at the origin of the room (0° azimuth). Eye movements were

monitored continuously by electrooculography in order to detect breaks of fixation between trials and during target presentation. Breaks of fixation (saccadic eye movements) were easily detected but rare (less than 1% of trials), and trials that contained errant saccades were repeated.

Continuous eye position and event times were sampled at a rate of 1 kHz by a realtime LabVIEW system (National Instruments, Austin, TX). Data analysis was performed in R ([www.r-project.org](http://www.r-project.org)) and Matlab (MathWorks Inc., Natick, MA).

### Stimulus Properties

Auditory targets consisted of 50 ms bursts of frozen broadband noise (0.2-20 kHz, 65 dB SPL, 1 ms  $\cos^2$  on/off ramps, equalized to have a flat spectrum). Between trials, continuous Gaussian white noise was presented at 65 dB SPL from two speakers mounted behind the screen at  $\pm 75^\circ$  horizontal,  $+20^\circ$  vertical, well outside of the range of target locations used in the experiment, in order to mask robotic sounds during target positioning. All auditory stimuli were generated by a TDT RX8 Multi I/O Processor (Tucker-Davis Technologies, Alachua, FL). Visual targets were 50 ms green laser dots, 3 mm (about  $0.1^\circ$  subtended angle) wide, that were projected onto the screen under program control of an X-Y mirror-galvanometer.

### Experimental Procedures

To elicit shifts in auditory spatial perception, subjects were exposed to trains of spatially disparate bimodal (AV) pairs (Figure 1b). AV trains typically consisted of 20 repetitions at fixed locations, in which visual and auditory targets were presented simultaneously but spatially displaced. Each bimodal AV pair was 50 ms in duration and repeated at 4 Hz, yielding a temporal sequence of 50 ms on, 200 ms off. AV presentations were positioned according to the auditory component, ranging  $\pm 30^\circ$  azimuth at intervals of  $3^\circ$ , and always at  $0^\circ$  elevation. The associated visual targets were displaced by  $+8^\circ$  (right) or  $-8^\circ$  (left) in azimuth for any given AV train, and the direction of AV disparity randomly varied from trial to trial.

To assure attention to both modalities during AV trains, one auditory or visual target was dropped from the AV disparity train and subjects were instructed to press a button as quickly as possible whenever they detected the dropped target. Dropped targets occurred with equal frequency in each modality, between the 3rd and the 19th repetition in the AV disparity train. Reaction times and false alarm rates were quantified to assess attentiveness during exposure trials. Detection of dropped targets was at or near 100% for most subjects, indicating that subjects sustained attention throughout the experiment. Two subjects in experiment 1 had poor detection rates for missing visual targets. However, both subjects had normal detection for missing auditory targets and their localization data did not differ from those of other subjects, and they were therefore included. On some trials in experiment 2 the AV disparity train only contained 1 repetition, so no targets were dropped. However, subjects could not predict these trials because they were randomly interleaved with trials that contained 20 repetitions, so vigilance was not threatened.

Subjects were asked to describe the spatial relationship between auditory and visual components of AV disparity trains at the end of their last experiment, and all subjects reported that they were unaware of the spatial disparity.

To measure shifts in auditory spatial perception due to visual capture and the ventriloquism aftereffect, subjects localized either the auditory component of AV disparity trains (*A within AV*) or auditory targets presented in isolation afterward (*A2 after AV*), respectively. Auditory targets presented afterward typically originated from the same location as the auditory target in the AV disparity train, except for a subset of trials in experiment 1. Localization was performed with an LED pointer connected to a cylindrical joystick that was illuminated 200 ms after the end of the delay period. This cued the subject to look and guide the pointer to the auditory target of interest (the auditory component of the AV train or the subsequent auditory target, depending upon trial type), and then register their response with a button press. Visual localization was not performed because the prior influencing visual capture does not appear to be influenced by the selective focusing of attention to either sensory modality individually or to both senses simultaneously (Odegaard et al., 2016).

To assess any enduring shifts in auditory spatial perception, auditory localization baselines were collected at the beginning and end of each experiment (pre- and post-AV trials). For each, subjects localized one repetition of the auditory target from each location used in these experiments. Auditory baselines in all experiments showed small changes between pre- and post-AV trials, but they proved idiosyncratic to each subject, indicating that these experiments did not elicit systematic and enduring changes in auditory spatial perception.

In all experiments, ocular fixation was monitored and carefully controlled during target presentation, as changes in eye position exert substantial shifts in auditory spatial perception (Cui and Razavi, 2010; Dobreva et al., 2012; Razavi et al., 2007). To minimize this bias, subjects always maintained center fixation during target presentation. To assist fixation between trials, a red laser beam was projected onto the center of the screen (0° azimuth and elevation) to provide a visual reference. The fixation laser was extinguished 100 ms prior to the onset of a trial, and remained off for the remainder of the trial. Subjects were instructed to maintain center fixation during target presentation despite the absence of this visual reference. Following target presentation, subjects were free to move their eyes to guide the LED laser pointer while localizing the target. Previous results from our lab demonstrate that eye movements do not bias localization of *remembered* auditory or visual targets (Dobreva et al., 2012).

The LED pointer was on between trials and when localizing remembered targets, but was extinguished during target presentation. Between trials, subjects pointed the LED pointer to the center visual fixation reference, and held this position (with the pointer turned off) throughout stimulus presentation to ensure that each localization response started from the same location.

## Experiment Protocols

The objective of *Experiment 1* was to determine the extent to which the ventriloquism aftereffect decays following exposure to AV trains. Ten subjects localized auditory targets

under two experimental conditions performed in separate blocks. In the first condition (*A within AV*), subjects localized the remembered auditory component of an AV train (20 repetitions at fixed disparity), using the laser pointer that appeared following a 1 s delay. Trials were presented to cover the  $\pm 30^\circ$  range of auditory space, as detailed above. This provided an estimate of visual capture. In a second condition (*A2 after AV*), a second auditory target (A2) was presented alone following the AV train, at the same location as the auditory component of the AV exposure but delayed by 1, 5, or 20 s. Additional trials with a delay of 1 s were presented in which A2 was displaced  $\pm 15^\circ$  in azimuth. This provided an estimate of the ventriloquism aftereffect at various delays and locations relative to the AV train.

The objective of **Experiment 2** was to determine; 1) whether visual capture or the ventriloquism aftereffect increases with the number of repetitions of AV trains; and 2) if visual capture decays even in the absence of additional targets, simply by waiting in dark silence. Ten subjects localized auditory targets under two conditions. In the first condition, subjects localized the remembered auditory component of AV trains (*A within AV*) after a 1 s delay, where the number of AV repetitions was either 1 or 20, presented in random sequence and across the range of azimuth. Further, in the case of 20 AV repetitions, trials were added in which localization was delayed by 10 s to assess the potential endurance of visual capture in the absence of additional input. In the second condition, subjects localized a second sound presented 1 s after AV trains (*A2 after AV*, at the same sound location), in which the number of AV presentations was either 1 or 20, as in condition 1. The two experimental conditions were alternated over 5 interleaved experimental blocks.

The objective of **Experiment 3** was to determine the extent to which visual capture endures in memory over time, even when subsequent auditory targets (A2) arise that might influence or compete with the memory of capture. Six subjects localized auditory targets under the same two conditions as above, related to what subjects are instructed to localize--*A within AV* or *A2 after AV* (1 s delay), and where AV trains always included 20 repetitions. However, a third and unique condition was added, in which AV is followed by A2, but subjects were instructed to localize either A2 (the usual *A2 after AV*), or the memory of *A within AV* while ignoring A2. The first condition was always performed first in a single experimental block, while the second and third conditions were performed in subsequent blocks, in random order counterbalanced across subjects.

## Results

We quantified visual capture and the ventriloquism aftereffect in response to  $8^\circ$  audio-visual disparities in azimuth. Our results demonstrate that both visual capture and the ventriloquism aftereffect occurred, but their effects on auditory spatial perception are distinct.

### Experiment 1: Spatial and Temporal Patterns of Auditory Shift Following an AV Disparity Train

The primary objective of this experiment was to characterize the temporal dynamics of perceptual shifts in auditory localization induced by the ventriloquism aftereffect. To do this,

we first quantified the visual capture effect induced by AV exposure; subjects simply localized the auditory component of the preceding AV train (*A within AV*). To address the ventriloquism aftereffect, subjects localized a second auditory target, presented following AV exposure and with varying time delays (*A2 after AV*). Further, we tested the spatial spread of the ventriloquism aftereffect by displacing A2 from the auditory component of AV trains (1 s delay only).

Exposure to AV trains with 8° of disparity strongly biased the localization of the auditory component of the AV stimulus toward the visual (Figure 2a; *A within AV*), consistent with visual capture (Bertelson and Radeau, 1981). We also observed a smaller shift when subjects localized a second auditory target presented after the AV train (Figure 2a; *A2 after AV*), which is consistent with previous measures of the ventriloquism aftereffect over brief time scales (Wozny and Shams, 2011). To quantify biases in auditory perception, we define *auditory shift* as the mean paired difference in response location between the left and right AV disparity conditions across all target positions, divided by two. Note that this definition of auditory shift is relative to individual responses across different disparity conditions rather than actual target location, because auditory localization is subject to an inherent bias toward the periphery (Razavi et al., 2007; Odegaard et al., 2015), which would alter the pattern of localization errors independent of visually induced biases. This peripheral bias in auditory localization is evident in Figure 2a as a non-flat slope of localization error as a function of target azimuth. Figures 2b and 2c show group trends and mean effects of AV disparity exposure on auditory localization as a function of A2 delay (b), as well as the azimuthal displacement of A2 (1 s delay only) from that within the AV train (c). There was a significant effect of delay on the magnitude of auditory shift (one-way repeated measures ANOVA,  $F_{2,18} = 15.2$ ,  $p < 0.001$ ), demonstrating that the ventriloquism aftereffect dissipates over time. Additionally, group mean auditory shifts were significantly different from zero at all but the 20 s delay condition (Bonferroni-corrected two-way t-test:  $p < 0.001$ ,  $p < 0.01$ , and  $p = 0.4$  for 1, 5, and 20 s;  $p < 0.001$  for both +15° and -15°). Thus, the ventriloquism aftereffect following 20 exposures to the AV train seems to dissipate within ~20 seconds, after which auditory localization effectively returns to baseline. Auditory shift decreased with displacement of A2 from the auditory component of AV trains by 15° in azimuth (one-way repeated measures ANOVA,  $F_{2,18} = 15.75$ ,  $p < 0.001$ ). This is in agreement with earlier studies on spatial dispersion of the ventriloquism aftereffect, in which exposure to short trains of AV disparities affected a limited range of auditory spatial perception (Bertelson et al., 2006; Kopco et al., 2009). Taken together, these results indicate that the ventriloquism aftereffect persists for seconds after exposure to disparity, and extends in space at least 15° in azimuth from the auditory component of the disparity.

## Experiment 2: Growth of Shift in Perceived Auditory Location with Repetition of AV Disparities

In this experiment we aimed to clarify whether visual capture and the ventriloquism aftereffect are sensitive to the duration of AV disparity. First, we quantified the effect of AV exposure duration on both phenomena by varying the number of presentations within AV trains for both *A within AV* and *A2 after AV* paradigms. Further, we also varied the delay



between AV trains (20 repetitions) and localization of *A within AV* to specifically determine whether visual capture decays over time, even in dark silence.

The auditory shift for *A within AV* trains did not increase with repetition (duration) of AV disparity and, it did not decay appreciably over a 10 s interval following AV presentation (Figure 3, *A within AV*). Increasing the number of repetitions of AV disparity from 1 to 20 had no significant effect on localizing *A within AV* (one-way repeated measures ANOVA,  $F_{1,9} = 2.49$ ,  $p = 0.15$ ). This is not surprising given earlier studies demonstrating near-maximal capture after a single exposure to disparity (Hairston et al., 2003; Wallace et al., 2004). Additionally, increasing the delay from 1 s to 10 s between the AV train and localization had no significant effect when localizing *A within AV* (one-way repeated measures ANOVA,  $F_{1,9} = 2.35$ ,  $p = 0.16$ ). This demonstrates that visual capture does not decay over time, even though the ventriloquism aftereffect does (experiment 1). In contrast, the auditory shift observed in *A2 after AV* conditions grew significantly with repetition of the AV exposure (Figure 3, *A2 after AV*, one-way repeated measures ANOVA,  $F_{1,9} = 19.89$ ,  $p = 0.002$ ). Thus, the ventriloquism aftereffect accumulates with experience, and also decays over time as demonstrated in experiment 1. There was no correlation between the magnitude of auditory shift due to visual capture and the ventriloquism aftereffect when both were measured after 20 repetitions of the AV disparity and given a 1 s response delay (simple linear regression,  $R^2 = 0.104$ ,  $p = 0.364$ ).

### Experiment 3: Interference of Subsequent Auditory Targets with Memory of an AV Disparity Train

This final experiment addressed whether auditory targets occurring after AV exposure (*A2*) directly affect visual capture within AV exposure. Specifically, we wondered whether the results in experiments 1 and 2 could be explained as a direct effect of *A2* targets disrupting the memory of visually captured auditory target location and thereby reducing the magnitude of auditory shift.

This experiment differs from those above in one important way. When a second auditory target (*A2*) was presented after AV trains, subjects were instructed to localize either *A2 after AV* or *A within AV* in different trials. In this latter case, note that while *A2* was presented, subjects had to ignore it in order to localize the sound within AV. Interestingly, adding *A2* when subjects were instructed to localize *A within AV* had no significant effect on auditory shift magnitude (Figure 4, compare the first and second conditions; one-way repeated measures ANOVA,  $F_{1,5} = 0.183$ ,  $p = 0.69$ ). The lack of effect indicates that the memory of visual capture is robust and can be held despite subsequent auditory events. In contrast, when subjects instead localized *A2 after AV*, the auditory shift was significantly reduced (Figure 4, compare second and third conditions; one-way repeated measures ANOVA,  $F_{1,5} = 22.57$ ,  $p < 0.01$ ), as a result of the ventriloquism aftereffect. This difference suggests that the ventriloquism aftereffect is not simply the result of visual capture being interrupted by a subsequent auditory event, and instead that the two phenomena are separate and have distinct influences on auditory spatial perception.

## Discussion

Our first finding is that the ventriloquism aftereffect dissipates over time and space. This is evident in the results of experiment 1, where the amount of auditory shift observed for auditory targets presented in isolation following an AV train decreased with increasing delay and spatial separation between the train and the localized target. This result demonstrates that the ventriloquism aftereffect induced by brief AV trains is local and transient, which is different from studies of the ventriloquism aftereffect after hundreds or thousands of repetitions of an AV disparity (Recanzone, 1998; Lewald, 2002; Wozny and Shams, 2011a). The difference between time scales suggests that either two separate mechanisms mediate short- and long-term ventriloquism aftereffects, or that the rate of temporal decay is proportional to the duration of exposure and that previous experiments did not measure post-exposure auditory perception for a long enough time to observe decay in auditory shift. The temporal decay of the ventriloquism aftereffect is also different from visual capture, which does not decay when held in memory over the same time scale, as shown in experiment 2.

Our second finding is that the ventriloquism aftereffect accumulates with repeated exposure to the same AV pair, but visual capture does not, as shown in experiment 2. This is unsurprising, because Wozny and Shams (2011) showed that repeated exposures to AV pairs with the same direction of disparity, even when those pairs were at different locations in space, caused a gradual buildup of shift in perceived auditory location. Additionally, Hairston et al. (2003) and Wallace et al. (2004) frequently demonstrated strong visual capture after just a single AV pair. Our current results match these previous studies, with the added benefit of demonstrating the distinction between visual capture and the ventriloquism aftereffect under matched conditions, with controlled eye position in the same subjects.

Our final finding is that the ventriloquism aftereffect cannot be explained as visual capture disrupted by subsequent auditory targets, but rather that the two mechanisms are different, occur simultaneously, and have distinct effects on auditory spatial perception. This point is first demonstrated in experiment 1, where despite little visual capture in two subjects, the ventriloquism aftereffect was still evident. The correlation analysis in experiment 2 reinforces this point and demonstrates that the auditory shift induced by visual capture and the ventriloquism aftereffect are uncorrelated under matched conditions. Experiment 3 confirms this finding by demonstrating that the magnitude of auditory shift induced by visual capture does not decrease when subsequent auditory targets are presented, indicating that visual capture and the ventriloquism aftereffect occur separately and simultaneously in response to the same stimulus.

In summary, visual capture does not accumulate with experience or dissipate over time, and is not disrupted by subsequent auditory targets. This suggests that the neural process that performs visual capture essentially acts as an immediate capture-and-hold sampler that binds related auditory and visual spatial cues into a common object which can be stored in memory for at least 10 seconds and is not altered by subsequent auditory experience. In comparison, the ventriloquism aftereffect accumulates with experience and dissipates over time and space. This indicates that the process that mediates the ventriloquism aftereffect acts as a leaky integrator that compensates for short-term changes in audio-visual

congruence to minimize cross-modal localization errors. At a broader level, results presented here indicate that although spatial disparities between auditory and visual targets drive both visual capture and the ventriloquism aftereffect, each phenomenon has distinct effects on auditory perception, and therefore they likely reflect separate neural mechanisms.

Our measurements of visual capture agree with previous findings. In both experiments, subjects were unaware of the spatial disparity between the targets, which is typical for audio-visual disparities smaller than 10-15° in azimuth (Bertelson and Radeau, 1981; Radeau and Bertelson, 1974). The majority of subjects showed large auditory shifts toward the visual target as a result of visual capture, with typical values around 6-7° of an 8° disparity. These large shifts are consistent with visual spatial dominance, due to vision's higher spatial precision (Battaglia et al., 2003). However, one subject showed very little visual capture in experiment 1 (this subject did not participate in experiments 2 or 3), and another that participated in both experiments 1 and 2 showed little capture in either experiment (this subject did not participate in experiment 3). A lack of capture in a portion of the subjects tested is expected because audio-visual binding is dependent on their prior expectation about the relationship between targets (Kording et al., 2007; Sato et al., 2007; Wozny et al., 2010), and subjects may vary in expectation that targets will originate from a common source. The magnitude of capture was also stable in subjects that participated in multiple experiments, which is expected because the prior expectation that targets originate from a common source tends to be consistent across experimental sessions (Odegaard and Shams, 2016). Altering the distribution of audio-visual disparities presented in an experiment can alter the expectation that targets originate from a common source (Van Wanrooij et al., 2010), but because we used a consistent magnitude of audio-visual disparity throughout all three experiments we do not believe this occurred. Additionally, visual capture occurred within single repetitions of an AV pair (experiment 2), which is in agreement with previous studies of capture (Hairston et al., 2003; Wallace et al., 2004; Wozny and Shams, 2011).

Our measures of the ventriloquism aftereffect also show similarities to previous work. Specifically, the ventriloquism aftereffect occurred even in subjects that showed little visual capture, indicating that capture does not govern the aftereffect (Wozny and Shams, 2011; Radeau and Bertelson, 1974; Radeau and Bertelson, 1978; Eramudugolla et al., 2011). The ventriloquism aftereffect was also evident after a single AV pair, in agreement with Wozny and Shams (2011). In our study, the magnitude of auditory shift after a single AV pair was about  $2^\circ/8^\circ = 25\%$ , which is larger than the approximately 10% shift observed by Wozny and Shams when the preceding AV pair was perceived to originate from a common source. However, their study used a broader range of audio-visual disparities, did not present the subsequent auditory target at the same location as in the AV pair, and the sample size in our study was relatively small, which could all account for the difference in shift magnitude.

Computational models of multisensory coding have demonstrated that optimal cue integration, which forms the basis for visual capture, can be readily implemented via physiologically plausible neural population coding (Ma et al., 2006). Additionally, both visual capture and the ventriloquism aftereffect can be implemented via synaptic projections between neural populations that encode visual and auditory spatial perception (Magosso et al., 2012), indicating that these phenomena could potentially arise from biologically

plausible projections between brain regions, rather than occurring within a dedicated neural structure. However, neurons that specifically respond to the spatial relationship between auditory and visual stimuli have been found in the inferior colliculus (for review, see Gruters and Groh, 2012), indicating that a dedicated neural population may underlie visual capture and the ventriloquism aftereffect. Similarly, neural activity in the auditory cortex of the planum temporale is associated with the occurrence of visual capture (Bonath et al., 2007; Bonath et al., 2014), activity in the posterior interparietal sulcus is associated with the integration or segregation of auditory and visual spatial targets (Rohe and Noppeney, 2015), and the ventriloquism aftereffect is also disrupted by lesions of the striate cortex (Passamonti et al., 2009), indicating that these cortical structures are also involved in the maintenance of audio-visual spatial congruence. Similarly, previous work on visual-vestibular integration analogous to visual capture has demonstrated that visual and vestibular heading cues are combined within specific multisensory neurons in dorsal medial superior temporal cortex (for review, see Angelaki et al., 2009), which supports the idea that multisensory integration and calibration in general may be implemented in populations of neurons specifically dedicated to this function, albeit in different regions of the brain. Future physiological experiments are required to identify neural mechanisms underlying the rapid temporal dynamics of the ventriloquism aftereffect described here. Additionally, the transience of the ventriloquism aftereffect following brief periods of audio-visual disparity shown here is distinct from the enduring shift known to occur after extensive exposure to hundreds or thousands of repetitions of an audio-visual disparity (Recanzone, 1998; Lewald, 2002; Wozny and Shams, 2011a). Therefore, our next step in understanding the behavioral temporal dynamics of the ventriloquism aftereffect is to characterize its short and long term effects on auditory spatial perception within the same experiment.

For natural behavior, visual capture and the ventriloquism aftereffect serve complementary functions. When an object provides salient auditory and visual cues simultaneously, visual capture during these cues acts to reduce cross-modal error and form a singular representation of that object. However, when tracking an object through a complex environment (for example, a lion moving through tall grass), localization cues from both senses are not always available. When visual information is unavailable (the lion is occluded by grass), the ventriloquism aftereffect maintains the same congruence that was provided by previous audio-visual cues, because the cross-modal error likely persists even when visual information is not currently available. However, if the previously observed cross-modal error only had limited evidence in its favor, it would be undesirable to permanently alter auditory spatial localization, because it could have been caused by an environmental artifact (for example, reverberations off of other objects in the environment) and not representative of an enduring change in sensory alignment. As a result, it would be best to track short-term changes in congruence after audio-visual disparities, but to have the effect of those disparities dissipate back to some baseline spatial map over time, as we have demonstrated here.

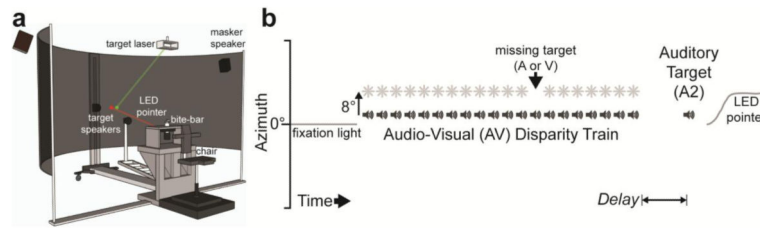
## Acknowledgements

We thank Martin Gira and Robert Schor for their technical assistance. Research was supported by NIDCD Grants P30 DC-05409 and T32 DC-009974-04 (Center for Navigation and Communication Sciences), NEI Grants P30-EY01319 and T32 EY-007125-25 (Center for Visual Science), and an endowment by the Schmitt Foundation.

## References

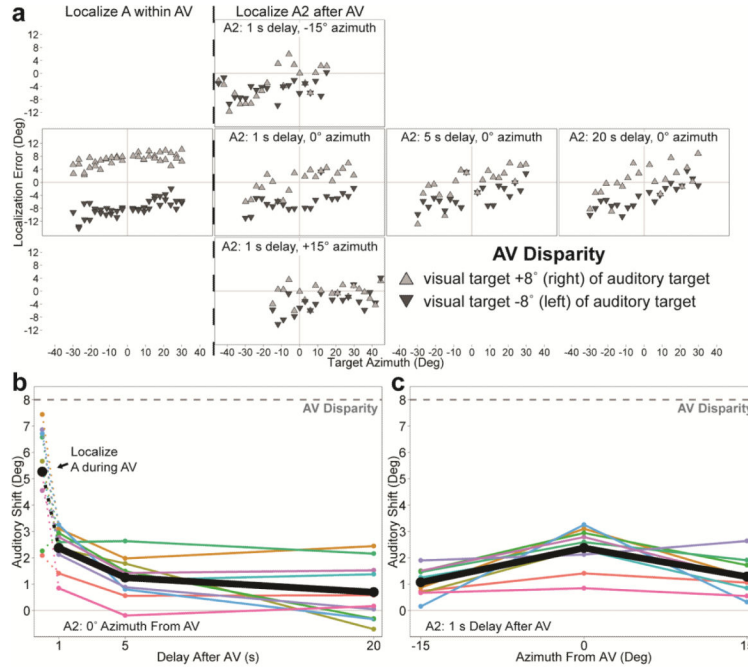
- Angelaki DE, Gu Y, DeAngelis GC. Multisensory integration. *Current Opinion in Neurobiology*. 2009; 19(4):452–458. [PubMed: 19616425]
- Battaglia PW, Jacobs RA, Aslin RN. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, Optics, image science, and vision*. 2003; 20(7):1391–1397.
- Bertelson P, Radeau M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception and Psychophysics*. 1981; 29(6):578–84. [PubMed: 7279586]
- Bertelson P, Frissen I, Vroomen J, de Gelder B. The aftereffects of ventriloquism: patterns of spatial generalization. *Perception and Psychophysics*. 2006; 68(3):428–436. [PubMed: 16900834]
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA. Neural Basis of the Ventriloquist Illusion. *Current Biology*. 2007; 17:1697–1703. [PubMed: 17884498]
- Bonath B, Noesselt T, Krauel K, Tyll S, Tempelmann C, Hillyard SA. Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. *NeuroImage*. 2014; 98:425–434. [PubMed: 24814210]
- Bruns P, Röder B. Sensory recalibration integrates information from the immediate and the cumulative past. *Sci Rep*. 2015; 5:12739. [PubMed: 26238089]
- Chen L, Vroomen J. Intersensory binding across space and time: a tutorial review. *Attention, Perception & Psychophysics*. 2013; 75(5):790–811.
- Cui QN, Bachus L, Knoth E, O'Neill WE, Paige GD. Eye position and cross-sensory learning both contribute to prism adaptation of auditory space. *Progress in Brain Research*. 2008; 171(08):265–270. [PubMed: 18718311]
- Cui Q, Razavi B. Perception of auditory, visual, and egocentric spatial alignment adapts differently to changes in eye position. *Journal of Neurophysiology*. 2010; 103(2):1020–35. [PubMed: 19846626]
- Dobrevá MS, O'Neill WE, Paige GD. Influence of age, spatial memory, and ocular fixation on localization of auditory, visual, and bimodal targets by human subjects. *Experimental Brain Research*. 2012; 223(4):441–455. [PubMed: 23076429]
- Eramudugolla R, Kamke MR, Soto-Faraco S, Mattingley JB. Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition*. 2011; 118(1):62–74. [PubMed: 20979992]
- Frissen I, Vroomen J, de Gelder B, Bertelson P. The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*. 2003; 113(3):315–327. [PubMed: 12835002]
- Frissen I, Vroomen J, de Gelder B, Bertelson P. The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychologica*. 2005; 118(1-2):93–100. [PubMed: 15627411]
- Frissen I, Vroomen J, de Gelder B. The aftereffects of ventriloquism: the time course of the visual recalibration of auditory localization. *Seeing and Perceiving*. 2012; 25(1):1–14. [PubMed: 22353565]
- Gruters KG, Groh JM. Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Frontiers in Neural Circuits*. Dec.2012 6:96. [PubMed: 23248584]
- Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA. Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*. 2003; 15(1):20–9. [PubMed: 12590840]
- Howard, I., Templeton, W. *Human Spatial Orientation*. Wiley; New York Jack: 1966.
- Jack CE, Thurlow WR. Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*. 1973; 37(3):967–979. [PubMed: 4764534]
- Kopco N, Lin IF, Shinn-Cunningham BG, Groh JM. Reference frame of the ventriloquism aftereffect. *The Journal of Neuroscience*. 2009; 29(44):13809–13814. [PubMed: 19889992]

- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS One*. 2007; 2(9):e943. [PubMed: 17895984]
- Lewald J. Rapid adaptation to auditory-visual spatial disparity. *Learning and Memory*. 2002; 9(5):268–278. [PubMed: 12359836]
- Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. *Nature Neuroscience*. 2006; 9(11):1432–8. [PubMed: 17057707]
- Magosso E, Cuppini C, Ursino M. A Neural Network Model of Ventriloquism Effect and Aftereffect. *PLoS ONE*. 2012; 7(8):e42503. [PubMed: 22880007]
- Mendonça, C., Escher, A., van de Par, S., Colonius, H. Predicting auditory space calibration from recent multisensory experience. *Experimental Brain Research*, 1983-1991. 2015.
- Odegaard B, Wozny DR, Shams L. Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLoS Computational Biology*. 2015; 11(12):1–23.
- Odegaard B, Shams L. The Brain’s Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological Science*. 2016; 27(4):583–591. [PubMed: 26944861]
- Odegaard B, Wozny DR, Shams L. The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*. 2016; 614:24–28. [PubMed: 26742638]
- Passamonti C, Frissen I, Lådavas E. Visual recalibration of auditory spatial perception: two separate neural circuits for perceptual learning. *The European Journal of Neuroscience*. 2009; 30(6):1141–1150. [PubMed: 19735289]
- Radeau, M., Bertelson, P. The Quarterly Journal of Experimental Psychology. Jun. 1974 The after-effects of ventriloquism; p. 37-41.2012
- Radeau M, Bertelson P. Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics*. 1978; 23(4):341–343. [PubMed: 748857]
- Razavi B, O’Neill WE, Paige GD. Auditory spatial perception dynamically realigns with changing eye position. *The Journal of Neuroscience*. 2007; 27(38):10,249–58.
- Recanzone GH. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95(3):869–875. [PubMed: 9448253]
- Rohe T, Noppeney U. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biology*. 2015; 13(2):1–18.
- Shams L, Beierholm UR. Causal inference in perception. *Trends in Cognitive Sciences*. 2010; 14(9): 425–432. [PubMed: 20705502]
- Sato Y, Toyoizumi T, Aihara K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Computation*. 2007; 19(12):3335–3355. [PubMed: 17970656]
- Thurlow WR, Jack CE. Certain determinants of the “ventriloquism effect”. *Perceptual and Motor Skills*. 1973; 36(3):1171–1184. [PubMed: 4711968]
- Van Wanrooij MM, Bremen P, Van Opstal JA. Acquired prior knowledge modulates audiovisual integration. *The European Journal of Neuroscience*. 2010; 31(10):1763–71. [PubMed: 20584180]
- Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. Unifying multisensory signals across time and space. *Experimental Brain Research*. 2004; 158(2):252–258. [PubMed: 15112119]
- Wozny DR, Beierholm UR, Shams L. Probability matching as a computational strategy used in perception. *PLoS Computational Biology*. 2010; 6(8)
- Wozny DR, Shams L. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of Neuroscience*. 2011; 31(12):4607–4612. [PubMed: 21430160]
- Wozny DR, Shams L. Computational characterization of visually induced auditory spatial adaptation. *Frontiers in Integrative Neuroscience*. Nov.2011 a 5:75. [PubMed: 22069383]
- Zwiers MP, Van Opstal aJ, Paige GD. Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience*. 2003; 6(2):175–81. [PubMed: 12524547]



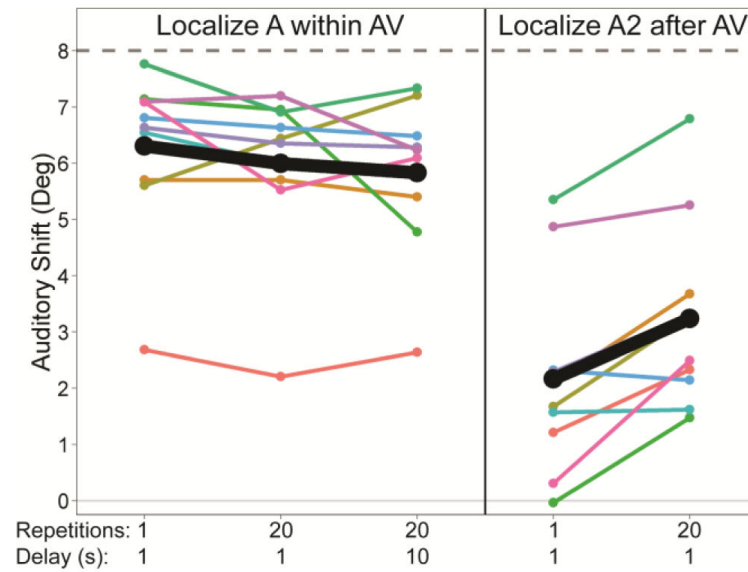
**Fig. 1.**

Experimental methods. **a**, Lab Schematic. **b**, In all experiments, subjects viewed a train of auditory-visual (AV) targets that were offset by  $\pm 8^\circ$  azimuth, and then localized either the auditory target presented within the AV train (to measure visual capture), or a second auditory target (A2) presented in isolation after a variable delay (to measure the ventriloquism aftereffect). To ensure consistent attention to AV trains, one auditory or visual target was absent from the train, and subjects were instructed to press a button whenever they detected the absent target. After target presentation (AV train, or AV train followed by A2), the LED pointer was illuminated and subjects localized the instructed target.

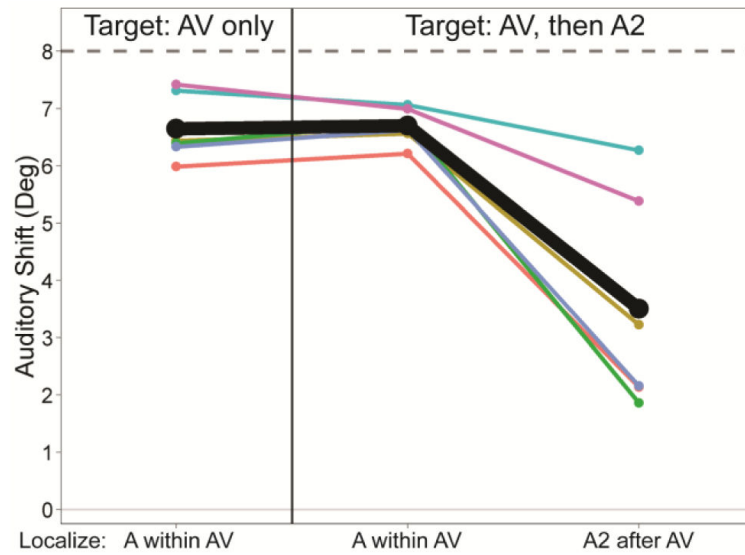


**Fig. 2.** Results from Experiment 1: Visual capture is apparent when subjects localized the auditory component of AV trains (*A within AV*) and the ventriloquism aftereffect is apparent when subjects localized the second auditory target presented after AV trains (*A2 after AV*). **a**, Auditory localization response errors from a representative subject are shown over a range of delays and azimuths with respect to the auditory component of AV disparity trains. Auditory spatial perception is predominantly shifted in the direction of the visual target, as demonstrated by the direction and magnitude of localization errors when the visual target was right (+8°) or left (-8°) of the auditory target. **b**, Visual capture is larger than the ventriloquism aftereffect, which dissipates over time and is no longer significant after 20 s. Colors represent individual subjects, while black indicates group means. **c**, The ventriloquism aftereffect on average extends to local space by at least 15°. Points represent auditory shift for A2 positions (1 s delay), all relative to the auditory component of AV disparity trains. Data at 0° azimuth are the same as the 1 s point in **b**.





**Fig. 3.** Results from Experiment 2: Visual capture when localizing *A within AV* is robust and unchanged with repetition, unlike the ventriloquism aftereffect when localizing *A2 after AV* task, where auditory shift is small following a single AV exposure but increases significantly with 20 repetitions. Further, visual capture persists after a 10 s delay of dark silence. Data were calculated and plotted as in Figures 2b and 2c.



**Fig. 4.** Results from Experiment 3: Visual capture when localizing *A within AV* is unaffected by subsequent auditory targets (*Localize A within AV*, with or without A2). Data were calculated and plotted as in previous figures.