JASA ARTICLE

Auditory-visual interactions in egocentric distance perception: Ventriloquism effect and aftereffect[a)]

Ľuboš Hládek,[1,b)] Aaron R. Seitz,[2,c)] and Norbert Kopčo[1]
[1]*Institute of Computer Science, P. J. Šafárik University, Jesenná 5, Košice, 040 01, Slovakia*
[2]*Department of Psychology, University of California Riverside, 900 University Avenue, Riverside, California 92521, USA*

ABSTRACT:
This study describes data on auditory-visual integration and visually-guided adaptation of auditory distance perception using the ventriloquism effect (VE) and ventriloquism aftereffect (VAE). In an experiment, participants judged egocentric distance of interleaved auditory or auditory-visual stimuli with the auditory component located from 0.7 to 2.04 m in front of listeners in a real reverberant environment. The visual component of auditory-visual stimuli was displaced 30% closer (V-closer), 30% farther (V-farther), or aligned (V-aligned) with respect to the auditory component. The VE and VAE were measured in auditory and auditory-visual trials, respectively. Both effects were approximately independent of target distance when expressed in logarithmic units. The VE strength, defined as a difference of V-misaligned and V-aligned response bias, was approximately 72% of the auditory-visual disparity regardless of the visual-displacement direction, while the VAE was stronger in the V-farther (44%) than the V-closer (31%) condition. The VAE persisted to post-adaptation auditory-only blocks of trials, although it was diminished. The rates of build-up/break-down of the VAE were asymmetrical, with slower adaptation in the V-closer condition. These results suggest that auditory-visual distance integration is independent of the direction of induced shift, while the recalibration is stronger and faster when evoked by more distant visual stimuli. © *2021 Acoustical Society of America*.
https://doi.org/10.1121/10.0007066
(Received 26 August 2020; revised 15 October 2021; accepted 19 October 2021; published online 15 November 2021)
[Editor: Laurie M. Heller] Pages: 3593–3607

I. INTRODUCTION

The human ability to localize sounds requires a lot of flexibility. For example, whenever we enter a new room, we need to adapt our mapping of acoustic localization cues to perceived locations to reflect the reverberation-related changes in those acoustic cues (Shinn-Cunningham *et al.*, 2005). Vision plays an important role in these calibration processes, causing both an immediate re-alignment of auditory spatial percepts to match the corresponding visual signals, and a long-lasting adaptation of the acoustic localization cue mapping based on the visual inputs. The immediate alignment of auditory percepts with visual signals is known as visual capture or the ventriloquism effect (VE) (Jack and Thurlow, 1973; Bertelson *et al.*, 2000; Alais and Burr, 2004; Recanzone, 2009). Persisting visually induced adaptation has been referred to as the ventriloquism aftereffect (VAE) (Kopčo *et al.*, 2009; Recanzone, 1998; Wozny and Shams, 2011). VE and VAE have been extensively studied for horizontal sound localization; however, fewer studies have researched how vision influences auditory distance perception (Anderson and Zahorik, 2014; Calcagno *et al.*, 2012; Cubick *et al.*, 2015; Gardner, 1968; Mendonça *et al.*, 2016; Mershon *et al.*, 1980; Mershon *et al.*, 1981; Rébillat *et al.*, 2012; Voss, 2016; Zahorik, 2001). The primary cues for auditory distance perception are sound level and direct-to-reverberant ratio. Visual distance perception relies upon binocular disparity and vergence cues, which also underlie auditory-visual interactions (*Tresilian* et al., 1999; Zahorik *et al.*, 2005). The main goal of the current study is to systematically examine the basic properties of the ventriloquism effect and aftereffect in distance.

One previously reported property of ventriloquism in distance is that the strength of the effect depends on whether it is induced by visual stimuli displaced closer than the auditory signal (a condition referred to here as V-closer) vs visual stimuli displaced farther away (a condition called here V-farther). For example, in anechoic space, VE creates a strong illusion of the sound coming from the nearest plausible visual target located closer than the sound source ("proximity image effect"; Gardner, 1968; Min and Mershon, 2005). Later, Mershon *et al.* (1980) found an effect similar to the proximity image effect that applied to visual targets located farther than the sound source. These effects are strong in anechoic rooms in part because very few intensity-independent distance cues are available for these sources (Brungart and Rabinowitz, 1999). In reverberant rooms, the visual dominance in distance dimension can be expected to be weakened, as intensity-independent

[a)]Portions of this work were presented at the 165th meeting of the Acoustical Society of America.
[b)]Current address: Audio Information Processing, Technical University of Munich, Munich, 80333, Germany. Electronic mail: lubos.hladek@tum.de, ORCID: 0000-0002-0870-7612.
[c)]ORCID: 0000-0003-4936-9303.

distance cues become available (Kopčo and Shinn-Cunningham, 2011). However, even in reverberation, auditory distance judgments are more accurate if the speakers are visible or if there are visual references in the room (Calcagno *et al.*, 2012; Zahorik, 2001). The current study examined whether there is an asymmetry in the strength of VE and VAE for the visual (V)-closer vs V-farther stimuli in a reverberant room.

Most previous ventriloquism studies examining ventriloquism have utilized one or only a few fixed visual stimulus locations combined with several auditory locations to examine the spatial dependence of the distance VE and/or VAE (Gardner, 1968; Mershon *et al.*, 1980; Zahorik, 2003). An important question that can elucidate the adaptive nature of the neural structures underlying ventriloquism is whether the auditory distance representation adapted is linear or non-linear (Bedford, 1993; Shinn-Cunningham *et al.*, 1998a). Typically, auditory distance studies analyze data on a logarithmic scale, as variance in distance judgments is approximately constant on this scale (Anderson and Zahorik, 2014; Kopčo *et al.*, 2012). The current study used a range of auditory-visual stimulus pairs with a fixed auditory-visual (AV) distance ratio to test whether adaptation by a constant amount in logarithmic units over a range of distances will induce constant VE and VAE shifts in logarithmic units. The relative strength of the induced VAE vs VE in both V-closer and V-farther conditions is examined to determine whether the size of aftereffects correlate with the size of VE and whether adaptation caused by ventriloquism is equally large for the two directions. Finally, the dynamics of the ventriloquism buildup and decay are addressed, as VE and VAE occur on different time scales and the rates of adaptation might also be different for the V-closer vs V-farther conditions.

To address these questions, a behavioral experiment was performed that investigated VE and VAE at intermediate egocentric distances from 0.7 to 2.04 m. The experiment was conducted in a small semi-reverberant room in which targets were located in a line in front of participants. The auditory (A) and V stimuli were presented on an array of light-emitting diodes (LEDs) densely covering the distances of 0.45 to 2.68 m, respectively. The ventriloquism effect was induced by presenting AV stimuli in which the A component (one of the 8 speakers) was paired with a V component (LED) positioned either 30% closer (V-closer) or 30% farther (V-farther). The size of V-component shift was fixed throughout the study so that the effects of displacement direction (V-closer vs V-farther) and stimulus distance could be examined. The ventriloquism aftereffect was examined in A-only trials interleaved with the AV trials, and by separate A-only runs performed before and after adaptation. The interleaved A-only trials were expected to show the immediate aftereffect caused by preceding AV stimulation on the timescale of seconds, while the shifts observed in the separate A-only runs were expected to reflect a more persistent ventriloquism adaptation on the timescale of minutes. Each session took approximately one hour and consisted of several experimental runs in which the direction of the V-stimulus displacement was fixed, combined with several pre-adaptation and post-adaptation control runs. While VE was expected to occur immediately upon the presentation of AV stimuli, VAE was expected to grow gradually as more and more AV stimuli were presented, and to decay gradually when the AV stimulus presentation stopped. The final analysis examined the VAE dynamics of adaptation across runs within a session, focusing on whether VAE strength differs for the V-closer and V-farther conditions, which would be another indication that the adaptation process underlying VAE is direction-dependent.

Many previous auditory distance and ventriloquism studies used virtual acoustics to present stimuli. Here, the main goal is to understand the relative effects of visual stimuli on auditory percepts, for which it is critical that the visual and auditory location percepts are veridical. Therefore, the study was performed in a real environment using loudspeakers and LEDs, to ascertain that the physical locations of the stimuli are unambiguous, and the percepts are as close to natural as possible. Of note, this naturalistic listening approach resulted in some artifacts, such as acoustic shadowing of speakers by each other. While this could have been avoided, e.g., by staggering the loudspeakers in azimuth or elevation, this design choice was made as staggering would result in additional azimuth or elevation cues that could help identify individual loudspeakers. Broadband noise stimuli were employed to provide strong localization cues such as the rich spectral cues which aid sound localization in distance and onset cues which are important for sound localization in rooms (Hartmann, 1983), while at the same time assuring that participants were unable to fixate on a specific spectral characteristic of the stimulus, as the noise was random and independent from trial to trial. LEDs were used as visual stimuli, as is commonly in ventriloquism literature (Bertelson *et al.*, 2006; Recanzone, 1998). They are small enough so that all of them were clearly visible in the dense array used here, and they have good temporal characteristics (quick onset and offset), which was important to keep the synchrony of the auditory-visual stimuli. To assess the possible effect of these experimental choices on baseline distance perception, Appendix A provides a comparison of the performance with current setup to available data from the literature.

II. METHODS

A. Participants

A group of 183 young adults participated in this study. All participants had normal or corrected-to-normal vision and normal hearing (by self-report). The experimental protocol was approved by the University of California, Riverside Human Research Review Board. All participants provided written informed consent.

https://doi.org/10.1121/10.0007066

B. Setup, stimuli, and procedures

The experiment was conducted in a small semi-reverberant acoustically treated room (T_{20} = [405, 444, 577, 552, 505, 411] ms in octave frequencies from 250 Hz to 8 kHz [Berzborn *et al.*, 2017; background noise 35 dB sound pressure level (SPL)] with internal dimensions of 2.6 m × 3.33 m × 3.05 m (H). Participants were seated on a barber chair with adjustable height and a headrest. The chair back faced the middle of one of the shorter walls in the room. The seated participant faced an array of nine uniformly spaced loudspeakers arranged along a straight path in front of the participant [Fig. 1(A)] at ear level. An acoustically transparent fabric covered the loudspeaker array, so the participant was unaware of the number and exact locations of the loudspeakers. Sound stimuli were delivered from 8 loudspeakers (Peerless 830984 Tymphany) mounted on stands. An additional loudspeaker placed closest to the listener was used to provide acoustic shadow so that every speaker was obstructed by at least one other speaker.

Visual stimuli were delivered from an array of forty-eight uniformly spaced LEDs (4.76 cm spacing) mounted on a wooden frame raised 7 cm (at the far end) to 10 cm (at the near end) above the loudspeakers. The array height slightly decreased with distance so that all LEDs were clearly visible to the participants. The same LED array was used to collect responses. The loudspeakers and the LED array were coupled with a multichannel digital signal processor controller (RX-8, Tucker Davis Technologies, Alachua, FL) and an 8-channel power amplifier (CROWN, Elkhart, IN). A trackball was used to control which light was lit and to collect the listener's responses. An experimental computer outside the experimental room controlled the experiment, stimulus presentation, and response collection. The experimental procedure, stimulus generation and data analysis were implemented in MATLAB (Mathworks, Natick, MA).

The auditory stimuli were 300-ms broadband noise bursts. On each trial, one noise token was randomly selected from a set of 50 pre-generated tokens, and presented from one of eight loudspeakers. The presentation sound level was held constant, which led to a natural decrease in sound level from 56 to 53 dBA with increasing target distance (measured, with reverberation, by a sound level meter Extech HD600, Flir Comercial System Inc., Elkhart, IN). A calibrated microphone positioned at the location of the listener's head and pointing towards the loudspeakers, connected to the recording and playback system, was used to record the stimuli for acoustic analysis. Frequency characteristics of the audio system and loudspeakers were estimated from the first 5.8 ms of impulse responses measured for each loudspeaker measured using the maximum length sequence technique (Rife and Vanderkooy, 1989; Vanderkooy, 1994). Figure 1(C) shows the octave-band spectrum of the direct portion of the responses. Responses are shifted systematically upwards as the distance of the speakers decreases. Colored traces in Fig. 1(E) show the same responses corrected for the distance-dependent level increase and after subtracting the across-distance mean at each frequency to assess speaker-to-speaker spectral differences and the effect of acoustic shadowing of the speakers. Responses vary mostly by up to 2 dB for frequencies up to 1 kHz, with no systematic effect of distance. At frequencies 2 kHz and higher, there is a systematic variation of up to 3 dB such that more distant speakers are attenuated relatively to closer ones (for the closest speaker at the highest frequency of 8 kHz, this deviation is more than 5 dB). Figure 1(E) also shows the spectra of the individual random-noise stimuli (gray traces). This figure shows that the stimulus-to-stimulus spectral variation for frequencies up to 1 kHz was on the order comparable with the speaker-to-speaker variation, which means that these two sources of variation were not separable for the listener. For frequencies of 2 kHz and higher, the across-speaker differences were systematic and larger than the stimulus-to-stimulus variation. However, for broadband stimulus distance perception, these frequencies are weighted much less than for the low frequencies of 500 Hz and less (Kopčo and Shinn-Cunningham, 2011). Although the differences could have provided some cues for distance perception, the study focuses on the effects of the visual stimulus and any such effects would be minimized by considering the performance in the baseline condition. For further reference on the distance perception in the baseline condition and comparison with the literature see Appendix A. Finally, Fig. 1(F) shows the A-weighed sound pressure level of the direct and reverberant portion of the stimuli, showing that the direct sound level decreased systematically with distance, while the reverberant portion was approximately constant at all distances.

The visual stimuli consisted of a 300-ms flash of light from one of the LEDs. The AV stimuli consisted of the auditory component (A-component, identical to the auditory stimulus) and the visual component (V-component, identical to visual stimulus) presented synchronously. The stimulus timing was controlled by a separate program running directly on the digital signal processor. The V-component of the AV stimuli was either aligned with the A-component (V-aligned), placed approximately 30% closer than the auditory component (V-closer), or approximately 30% farther than the auditory component (V-farther). The exact relative placement of the V-component vs A-component deviated slightly from 30% because of the discrete positions of the LEDs [Fig. 1(B)]. Also note that, while the 30% shift results in a constant shift as a function of target distance in log units when the V-closer and V-farther positions are considered separately, it corresponds to a larger absolute logarithmic value in the V-closer direction than in the V-farther direction. Specifically, 70% of target distance is –0.36 log(cm) while 130% distance is 0.26 [compare the left-hand vs right-hand axis labels in Fig. 1(B)], which makes it more complicated to directly compare the effects of V-closer and V-farther adaptors in logarithmic units. A correction was applied to address this disparity, as described in the results section.

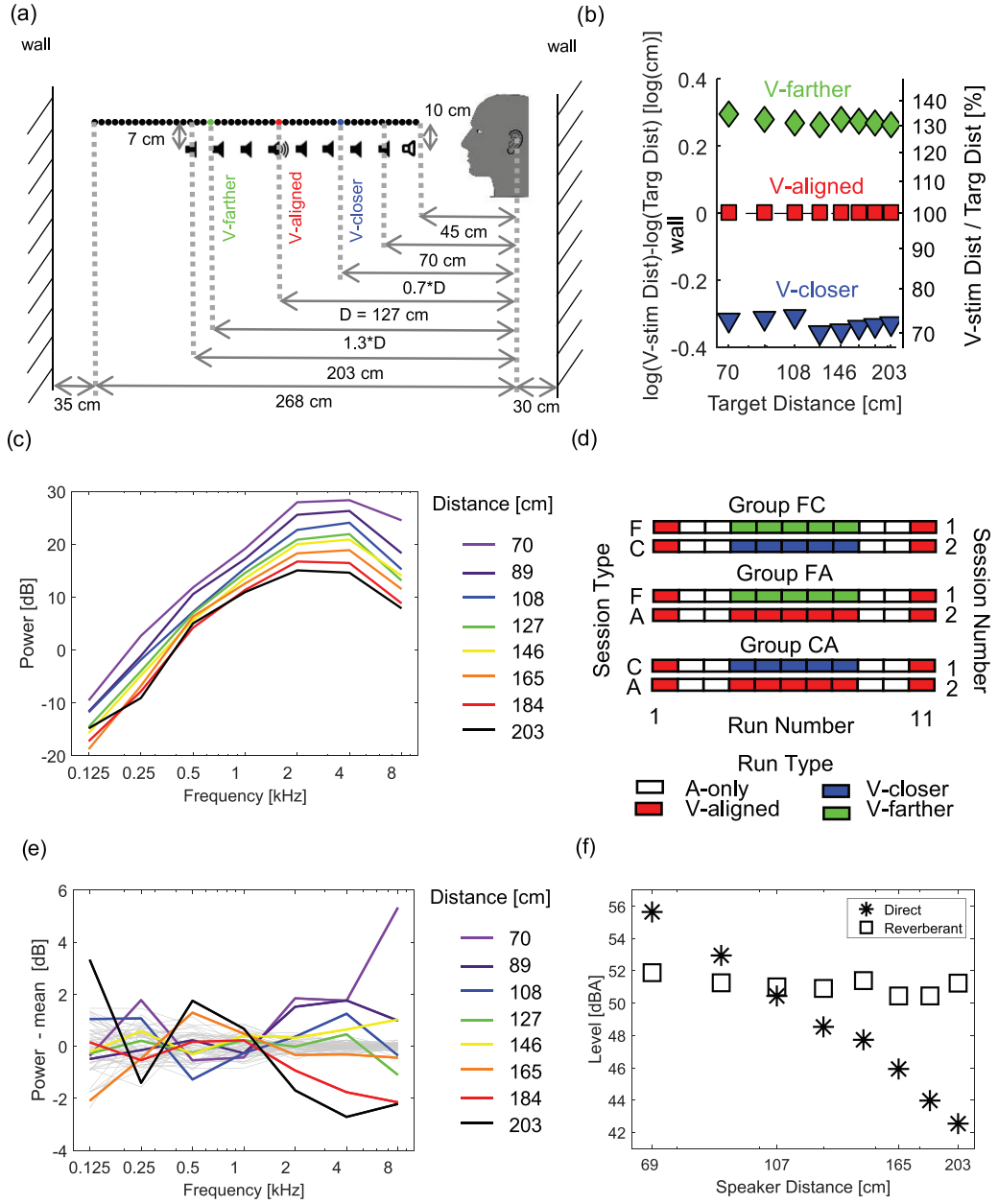

FIG. 1. (Color online) Setup, stimuli, and groups. (A) Setup and locations of the subject, loudspeakers, and LEDs in the room. Loudspeakers were spaced linearly, and the nearest loudspeaker (50 cm) was not used. An example AV stimulus pair is also shown, with the loudspeaker at $D = 127$ cm (shown as the active loudspeaker) and the corresponding V-farther LED (green), V-aligned LED (red), and V-closer LED (blue). (B) Actual physical locations of the V-components of AV stimuli used to approximate the 30% AV separation in the V-farther and V-closer conditions. The V-component distance is shown re. the A-component (on a log scale) as a function of the target location. Right-hand side axis shows the ratio of distances in per cents. The x axis was log transformed and this transformation is used throughout the paper. (C) Frequency response of the audio delivery system measured for each loudspeaker in 1-octave bands at the location of the listener head (without reverberation). (D) Schematic of the experimental session structure for different participant groups. Each participant was assigned into one of three groups and participated in two sessions of eleven runs. Each row represents a series of runs performed by the participant from a given group from a given session. Session types F, A, and C differed by the type of AV stimulus (V-farther, V-aligned, and V-closer) used in the adaptation runs 4-8. Session ordering was counterbalanced across participants within each group. (E) Thick colored lines show, for each loudspeaker, the deviation of frequency response of that loudspeaker (from panel C) with respect to the across-loudspeaker mean, corrected for the distance-dependent overall level change. Thin gray lines show frequency response of each stimulus with respect to the across-stimulus mean computed in 1-octave bands. (F) A-weighted sound pressure levels of the direct and reverberant portions of the experimental signals. Critical distance is 103 cm.

Each participant participated in two experimental sessions, each consisting of 11 runs of 64 trials. Each trial started with the presentation of an A-only or AV-stimulus, followed by the participant's response. The participant's task was to indicate the perceived egocentric distance of the stimulus. When the AV stimulus was presented, participants were instructed to respond to the A component and ignore the V component. To collect participant responses, a random LED was activated after the presentation of the stimulus, and the participant then moved the active LED towards a perceived position using the trackball. Only one LED was active at any given time. The participant submitted their

https://doi.org/10.1121/10.0007066

response with a button-click on the trackball (Wozny and Shams, 2011). No immediate feedback was provided, and after 500 ms inter-trial interval, the next trial was initiated.

There were two types of runs. The A-only runs consisted of auditory-only trials. The AV runs consisted of A-only and AV trials pseudorandomly interleaved with the ratio of 3:1 (75% AV trials, 25% A trials). In each run, each of the eight target loudspeakers was randomly chosen to present stimulus on two auditory trials and six AV trials. The relative displacement of the visual component was fixed within AV runs. The AV runs were of three types, V-closer, V-farther, or V-aligned. In each A-only run, each target loudspeaker was randomly chosen eight times.

There were three types of sessions, F, C, and A, differing by what type of shift was being induced [V-farther, V-closer, or V-aligned, respectively; Fig. 1(D)]. Each session, independent of its type, started with a pre-adaptation baseline period consisting of one V-aligned AV run followed by two A-only runs. The AV trials in run 1 were included both to minimize possible pre-adaptation biases that participants might have in their auditory-only distance responses and to establish a pre-adaptation reference. The sessions differed only in the adaptation period consisting of five AV runs (runs 4–8) with the AV disparity fixed within the session, which determined the type of session (C, F, or A). The post-adaptation period consisted of two A-only runs to assess the persistence of the induced adaptation and one V-aligned AV run to minimize any carry-over effects of adaptation on participant. A whole session took approximately 1 h to complete. There were 30-s breaks between the runs within a session. The participant received feedback about their progress during these breaks *via* voice commands played by one of the loudspeakers in the array.

Participants performed two sessions that were conducted on separate days. Each participant was randomly assigned to one of three experimental groups [Fig. 1(D)] differing by the conditions they performed in the sessions. Group CF (34 participants) performed the F and C sessions, Group FA (40 participants) the F and A sessions, and Group CA (40 participants) the C and A sessions. The order of sessions was counterbalanced across participants within a group.

A separate preliminary calibration measurement was performed to establish if any biases in responses might be due to the experimental setup and, in particular, the response method based on a manually controlled visual pointer and/or due to biases in visual distance perception. In the calibration measurement, a group of new 69 participants judged the egocentric distance of visual stimuli, using procedures identical to those described above, except for the following two differences. First, the calibration session had only one run of 80 trials. Second, on each trial, a visual stimulus was presented from one randomly chosen LED in the array of 48 LEDs such that each LED was chosen at least once and not more than twice.

C. Data analysis

The data from the sessions of the same type were pooled across the subject groups; e.g., the V-closer condition data were obtained from the C sessions of the participants from Group CF and Group CA. All responses were converted to the log scale (Anderson and Zahorik, 2014; Kopčo *et al.*, 2012) before any manipulations. When possible, percent scale is also plotted in graphs to provide a more intuitive description of the data. However, note that percent scales cannot be used to describe the error bars as the scale is non-linear. Unless specified otherwise, all results graphs show the across-subject mean and standard errors of the mean (SEM). Analysis of variance (ANOVA) with within-subject factor of distance (1–8) and between-subject factor of session type (F, C, A) assessed statistical significance of effects. The statistical software CLEAVE (Herron, 2005) was used for the ANOVAs and the reported p values were corrected by using the Greenhouse-Geisser epsilon.

III. RESULTS

The main goal of this study was to examine the effect of visual stimulation on auditory distance perception. To separate this effect from possible biases due to the experimental methods used here (like sensitivity to the exact placement of the speakers) and/or inherent biases in auditory distance perception, data from adaptation runs were analyzed relative to the pre-adaptation baseline runs 1–3. This section first presents the baseline data and a detailed comparison with other auditory distance studies is provided in Appendix A. Then, effects of congruent and incongruent visual stimuli on VE and VAE are presented as a function of target distance for the adaptation runs 4–8 and the post-adaptation runs 9–10. Finally, dynamics of the VE and VAE build-up and breakdown are analyzed as a function individual runs within a session for data collapsed across distance.

A. Baseline

Figure 2 shows baseline reported distance (upper row) and localization bias (lower row) for the Visual-only stimuli [Figs. 2(A) and 2(D)], V-aligned AV stimuli [Figs. 2(B) and 2(E)], and Auditory-only stimuli [Figs. 2(C) and 2(F)]. The visual baseline [Figs. 2(A) and 2(D)] was measured to validate the response method used here. The Visual-only responses were very accurate in the range in which the auditory targets were presented (70–203 cm), with only slight undershooting (by less than 5%) at the largest distances. However, this undershooting became more pronounced at larger distances, and reached 10% at the distance of 250 cm. It is not clear whether this bias is caused by actual biases of the visual distance perception or by some artifact of the response method. However, it is likely that this bias only influenced measurements for the most distant targets in the V-farther condition, for which the visual components were presented at distances larger than 2 m. To minimize the possible effect of this response method bias on the VA and VAE data, the data were corrected by adding the inverse of the response method bias to the baseline-corrected perceived location of auditory and AV targets (see below). Since this correction only affected responses larger than 2 m,

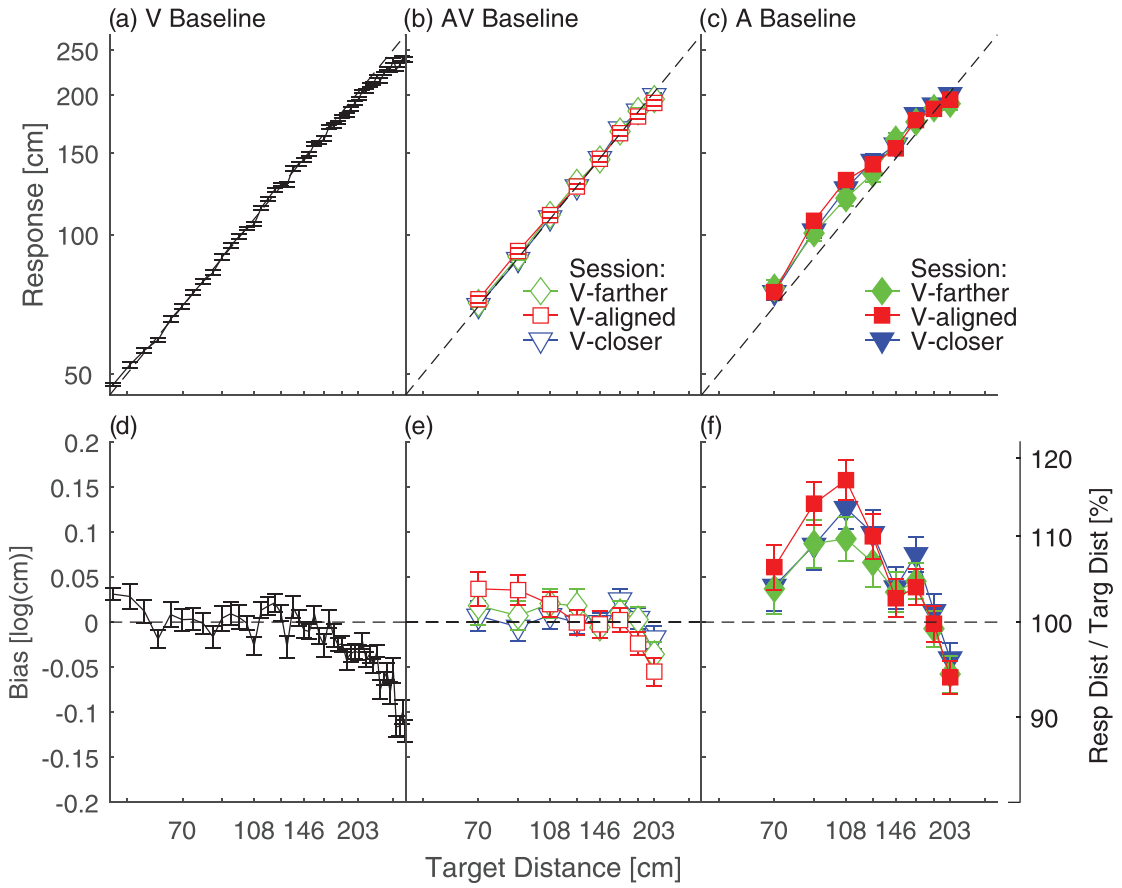

FIG. 2. (Color online) Baseline performance as a function of target distance. Upper panels: Across-subject geometric mean of the response distance as a function of actual target distance. Lower panels: across-subject bias in log of response *re*. log of actual target distance. (A) and (D), Visual-only stimulus baseline measured in a calibration experiment. (B) and (E), Auditory-visual baseline measured in run 1 of each session. (C) and (F), Auditory baseline from runs 1–3 in each session. Different symbols in (B), (C), (E), and (F) represent different types of sessions. The right-hand *y* axis of panels (D)–(F) show the ratio of the response and target distance in percent.

additional statistical analysis is also provided, where relevant, with the three most distant speakers excluded (considering speakers below 1.5 m), to assure that the reported results are not affected by the correction (and Appendix B describes the results of this analysis without applying the V-response correction).

The AV baseline [Figs. 2(B) and 2(E)] was measured in run 1 which used the V-aligned AV stimuli. Very little bias was observed (less than +/–5%), with an undershooting trend at the largest distances, consistent with the bias observed in the Visual-only calibration measurement [Fig. 2(A)]. Importantly, there were no large differences between the three conditions (shown by different symbols), indicating that the groups were similar in their AV baseline performance.

Figures 2(C) and 2(F) show the auditory baseline obtained from auditory responses in runs 1–3. Again, there were no systematic differences between subject groups (shown by different symbols). However, participants tended to overestimate the auditory targets at most distances. The overestimation was largest (10%–20%) at distances around 100 cm, and gradually decreased at larger distances, flipping to underestimation by approximately 5% at 200 cm [this underestimation, again, is likely caused by the bias observed in Fig. 2(A)]. This result is consistent with the auditory horizon observed in previous studies reporting that distance estimates are overestimated for distances less than approximately 1–3 m and underestimated for larger distances (Zahorik, 2001). In the current study, the location of the horizon appears to be at around 200 cm. Inconsistent with previous studies is the fact that the bias was reduced at the shortest distances. Even though a slight similar trend is observed in (Kopčo and Shinn-Cunningham, 2011), most previous studies performed in virtual environments show that overestimation grows as the sources approach the listener (see Appendix A, which compares the current baseline to an additional Auditory-only baseline measurement and to Anderson and Zahorik, 2014; Kopčo and Shinn-Cunningham, 2011; Rébillat *et al.*, 2012; Zahorik *et al.*, 2005; Zahorik and Wightman, 2001). Overall, this reduced bias suggests that, in the real environment and with the AV-aligned stimuli interleaved in the initial run, as used here, participants are more accurate at judging the distance of nearby sources than in virtual environments, even if the loudspeakers shadow each other.

Given the large overestimation observed in the previous auditory distance studies, it is not surprising that these small

https://doi.org/10.1121/10.0007066

(up to 20%) biases in the A-only responses are observed, even though in run 1 the A-only stimuli were interleaved with V-aligned AV stimuli for which no bias was observed (panels B and E). To account for this, the analysis of the VE and the VAE in the following sections (III B–III E) plots data relative to this A-only pre-adaptation baseline. And, since this study focuses on relative effects of the V-farther and V-closer stimuli referenced to the V-aligned condition, it is assumed that the baseline biases are not differentially affecting these relative comparisons.

B. Ventriloquism effect and aftereffect

The upper panels of Fig. 3 show the raw reported distances as a function of target distance, separately for the three types of sessions (represented by different symbols and colors). Figure 3(A) shows responses to AV stimuli during the adaptation runs 4–8, Fig. 3(B) shows responses to the A-only stimuli during the adaptation runs 4–8, and Fig. 3(C) shows responses to the A-only stimuli in the A-only post-adaptation runs 9–10. The main trend shown in these panels is that estimates grow with the target distance. To focus on the effect of the visual stimuli instead of the overall effect of distance, each panel in the middle row [Figs. 3(D)–3(F)] shows data from the respective upper panel, plotted relative to the response distance baselines [from Fig. 2(B) and 2(C)], and corrected for the response method biases by adding the inverse of the response method bias [computed from data in Fig. 2(A)].

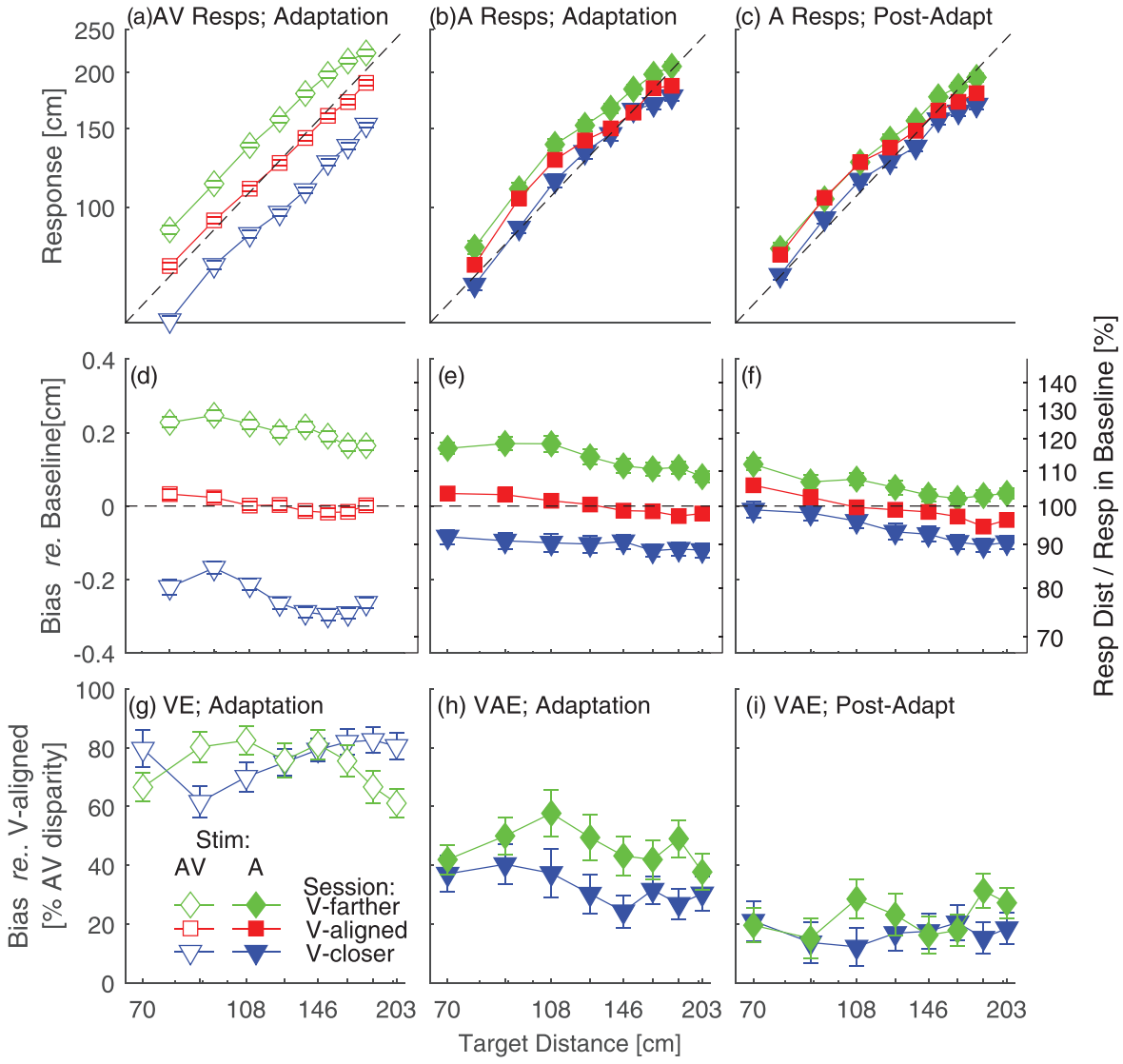

FIG. 3. (Color online) Performance in adaptation and post-adaptation runs. (A)–(C), Across-subject geometric mean of response distance as a function of target distance. (D)–(F), Across-subject mean of the response distance (from respective A–C) *re*. Baseline-run response distance (from Fig. 2) plotted as a function of target distance. (G)–(I), Across-subject mean of the V-closer (V-farther) data from the respective middle-row panel (D)–(F), plotted relative to the corresponding V-aligned data (from the same middle-row panel) and normalized by the physical disparity between the visual component and the target [from Fig. 1(D)]. The three types of sessions are shown by different symbols. The left-hand and middle columns show data averaged across adaptation runs 4–8 for AV targets (left-hand column) and A-only targets (middle column). The right-hand column (C, F, and I) show the A-only data averaged across post-adaptation runs 9 and 10. The right-hand *y* axis of (D)–(F) show the ratio of the response distance and the baseline-run response distance (in per cent). Data in (D)–(I) were corrected for the response method bias shown in Fig. 2(D).

Finally, the lower panels of Fig. 3 show the magnitudes of the ventriloquism effect [VE; Fig. 3(G)], immediate after-effect [VAE; Fig. 3(H)], and persistent aftereffect [persistent VAE; Fig. 3(I)] derived from the data in the corresponding middle panels of the figure. Specifically, to compute the effect of a visual component placed closer (farther) than the auditory component in the lower panels, the differences between the corresponding V-closer (V-farther) and V-aligned data are taken from the middle panel and scaled by the physical disparities between the visual and auditory components [from Fig. 1(B)].

AV responses in the V-aligned condition [red squares in Fig. 3(D)] are approximately at 0 log(cm), slightly overshooting near targets and undershooting distant targets. The V-farther (green diamonds) and V-closer (blue triangles) data are both shifted in the direction of the visual component, confirming the existence of the VE in the distance dimension. Also, their dependence on target distance is similar to that for the V-aligned data (slight downward slope). On average V-farther responses are shifted up by 0.2 log(cm) and V-closer responses are shifted down by 0.25 log(cm).

Similar to the AV-responses in panel D, the auditory-only responses during adaptation runs [Fig. 3(E)] show a slight compression (downward slope) in all conditions. There is also a clear effect of the interleaved AV stimuli, such that both V-closer (blue triangles) and V-farther (green diamonds) stimuli shift the auditory-only responses in the direction of the visual component. The effect is approximately linear on the logarithmic scale (the blue and green lines are parallel with the red line); however, it is weaker than in the AV responses [Fig. 3(D)], in particular for the V-closer stimuli (compare the filled blue triangles of Fig. 3(E) to the open blue triangles of Fig. 3(E)].

In post-adaptation runs [Fig. 3(F)], the observed trends in the A responses are similar to those during adaptation [Fig. 3(E)], but the magnitude of the effect of V-closer and V-farther adaptation is smaller than that in Fig. 3(E), and approximately equal for all targets independent of their location or the direction of the visual component shift.

The downward sloping red lines (as well as the green and blue lines) in Figs. 3(D)–3(F) indicate that responses tended to be more compressed during the adaptation and post-adaptation relative to the baseline. While it is not clear what drove this trend, it might be due to within-run accommodation of subjects to the experimental procedure, or due to decreased attentiveness or increased tiredness during the experimental session, which, on average, is expected to result in more bias towards the middle of the response range. Also, note that the bias is slightly larger in the post-adaptation runs 9–10 [Fig. 3(F)] than in the adaptation runs 4–8 [Fig. 3(E)]. Importantly, this bias can be approximately eliminated when comparing the effect of displaced V-stimuli (V-closer or V-farther) to the V-aligned reference, as it tends to affect them equally.

In summary, all of the results shown in Figs. 3(D)–3(F) show a clear ventriloquism effect and aftereffect in distance perception. The effect appears approximately constant as a function of target distance and depends in strength on the direction of the induced shift (closer vs farther). Sections III B through III D provide a detailed analysis of these results.

C. Ventriloquism effect

Two trends in Fig. 3(D) suggest that additional corrections need to be applied before the effects of interest can be properly evaluated. First, the AV response biases in Fig. 3(D) are overall larger for the V-closer stimuli than the V-farther stimuli (relative to the V-aligned stimuli; compare the blue triangles and green diamonds to the red squares). However, this asymmetry is likely caused by the fact that a 30% shift of the visual component corresponds to a larger disparity in logarithmic units for the V-closer than V-farther condition. Second, while the V-farther data (green diamonds) are relatively smooth, the V-closer graph (blue triangles) has a peak at the 89-cm target and a trough at the 165-cm target. This non-uniformity might be due to the slight variability in the size of the physical disparity between the V and A components, which was not always exactly 30% [Fig. 1(B)]. To account for these artifacts, as well as for the compressive bias observed even in the V-aligned condition, the V-closer and V-farther data from Fig. 3(A) were re-plotted in Fig. 3(G) relative to the across-subject average of the V-aligned data and scaled by the size of the physical AV disparity. Thus, to show 7, 1022 strength of the ventriloquism effect, Fig. 3(G) plots the size of the shift in the perceived location as a proportion of the shift in the physical displacement of the V-component which induced it.

VE data from Fig. 3(G) were subjected to a two-way mixed ANOVA with factors of target location and condition, which showed that the VE is approximately equally strong in the V-closer and V-farther conditions [main effect of condition: $F(1,146) = 0.29$, n.s.; for targets<1.5 m, i.e., for targets unaffected by the V-response correction: $F(1,146) = 0.63$, n.s.]. While there was considerable variation in the size of the effect as a function of target distance in both the V-closer and V-farther graphs [interaction of condition and target distance: $F(7, 1022) = 5.97$; $p < 0.0001$; for targets < 1.5 m: $F(4,584) = 4.72$, $p = 0.0016$], there was no clear systematic trend of increasing or decreasing effect size with target distance. Moreover, this variation is relatively small compared to the overall VE and it is likely driven by noise, e.g., due to factors like the limited accuracy in the estimation of the V-aligned responses. Specifically, the observed pattern of maxima and minima in the V-closer data is exactly mirroring that in the V-farther data (i.e., the maxima in the green line are aligned with the minima in the blue line, and *vice versa*), which would be expected if the effect is mostly driven by noise in the V-aligned data estimation, as that estimate gets subtracted from the V-farther data and added to the V-closer data. Therefore, the dependence of results on target distance, while significant, was not further examined here.

https://doi.org/10.1121/10.0007066

In summary, the effect of spatially displaced visual stimulus on the perceived distance of a simultaneously presented auditory stimulus is approximately constant on a logarithmic scale at 72% of the visual stimulus displacement when the visual component is shifted by a constant 30% *re.* auditory component over a range of distances from 0.7 to 2 m. This effect is independent of whether the visual component is located closer or farther than the auditory component (V-closer vs V-farther) and it does not show any clear trend of systematic increase or decrease with target distance (values ranging from 60% to 80%). Thus, there is no evidence that the ventriloquism effect measured in a reverberant room is stronger for the V-closer direction, as suggested for anechoic space by Gardner (1968). And, the data are consistent with the suggestion that auditory distance is represented in logarithmic units, as the observed VE was approximately constant at 72% of the imposed 30% AV disparity in all the VE measurements.

D. Ventriloquism aftereffect

The ventriloquism aftereffect (VAE) was examined on two timescales. First, the immediate VAE was assessed for the A-only trials randomly interleaved with the AV trials during the adaptation runs [Fig. 3(H)]. This effect reflects adaptation on the timescale of seconds and tens of seconds. Second, persistent VAE was assessed by two post-adaptation A-only trial runs, reflecting adaptation on the timescale of minutes [Fig. 3(I)].

To evaluate the immediate VAE, data from Fig. 3(E) were replotted in panel H relative to the V-aligned data and scaled by the size of the physical AV disparity, using the same procedure and rationale as discussed above for the VE. These data were subjected to a mixed two-way ANOVA with the factors of target location and condition which showed that VAE was independent of the target location [$F(7, 1022) = 1.81$; $p = 0.08$; for targets < 1.5 m: $F(4,584) = 2.14$, n.s.], but that it was clearly stronger for the V-farther condition (approximately 44% of the AV offset) than for the V-closer condition [approximately 31% of the AV offset; $F(1,146) = 5.48$; $p = 0.02$; for targets < 1.5 m: $F(1,146) = 4.40$, $p = 0.037$].

To evaluate the persistent VAE, the data from panel F were again rescaled and replotted in Fig. 3(I). The observed patterns are also similar to the immediate VAE, except that the effect is smaller (on average, 19% of the physical AV displacement). ANOVA performed on these data did not show any significant effect or interaction ($p > 0.1$; same for targets < 1.5 m). However, the green diamond data tend to fall above the blue triangle data in Fig. 3(I), showing that there still is a trend for the aftereffect to be stronger in the V-farther condition.

In summary, the ventriloquism aftereffect induced by the AV disparity is similar to the ventriloquism effect in that it is constant as a function of distance. However, contrary to the VE, the VAE is asymmetrical, being considerably stronger in the V-farther direction. Specifically, when the immediate VAE is expressed as a proportion of VE, the average across-distances is 63% in the V-farther direction and only 43% in the V-closer direction. Similarly, when the persistent VAE is expressed as a portion of VE, it is on average 29% and 22% in the V-farther and V-closer conditions, respectively (see Fig. 6 in Appendix B for the data without the correction for the visual baseline). There are at least two possible explanations of this farther-closer asymmetry. Either the neural distance representation adapted in the ventriloquism aftereffect is more flexible in one direction than the other, or the visual signals that cause the adaptation are stronger or more salient in the V-farther direction. Importantly, the fact that VAE observed here is constant as a function of target distance for both directions is consistent with the assumption that the neural representation of auditory distance is adaptable in logarithmic units, at least over the range of distances and for the AV disparity examined in the current study.

Finally, no adaptation was observed in the V-aligned data, even though the V-aligned A-only baseline was shifted by up to 10%–20% re. the AV baseline in Figs. 2(F) vs 2(E). It would be expected that, if the baseline misalignment was perceptual, the shift would be corrected by the V-aligned signals during the VAE adaptation due to multisensory enhancement (Bruns *et al.*, 2020). Since no correction occurred, the disparity in baselines is more likely caused by differences in response strategies than in perception, e.g., since the same LED array was used to present the V-stimuli and collect the responses, the responses might be more accurate for stimuli with V-components than for auditory-only stimuli, as the subjects might be able to directly compare the perceived location of the visual-stimulus and the response LED.

E. Build-up and break-down of the ventriloquism effect and aftereffect

Although the observed effects of target distance showed significance in the analysis presented in the preceding section, the magnitude of the variation between individual target locations was relatively small compared to the magnitude of the overall effect. Therefore, the dynamics of the build-up and break-down of the visually induced adaptation were analyzed after collapsing the data across target location (while keeping the data from each experimental run separate). It was expected that the VE build-up and break-down would be immediate while the VAE build-up and break-down would have slower dynamics. Also, it was expected that, given the difference in the strength of the VAE for the V-farther vs the V-closer condition, there might be a difference in the rate at which the VAE builds up for the two directions.

Figure 4(A) shows the localization biases in the AV (open symbols) and A-only (filled symbols) trials separately for each run as a function of the run number, referenced to the pre-adaptation runs 1–3. Figure 4(B) shows the magnitudes of the VE and VAE computed by referencing the data from panel A to the V-aligned condition and scaling them

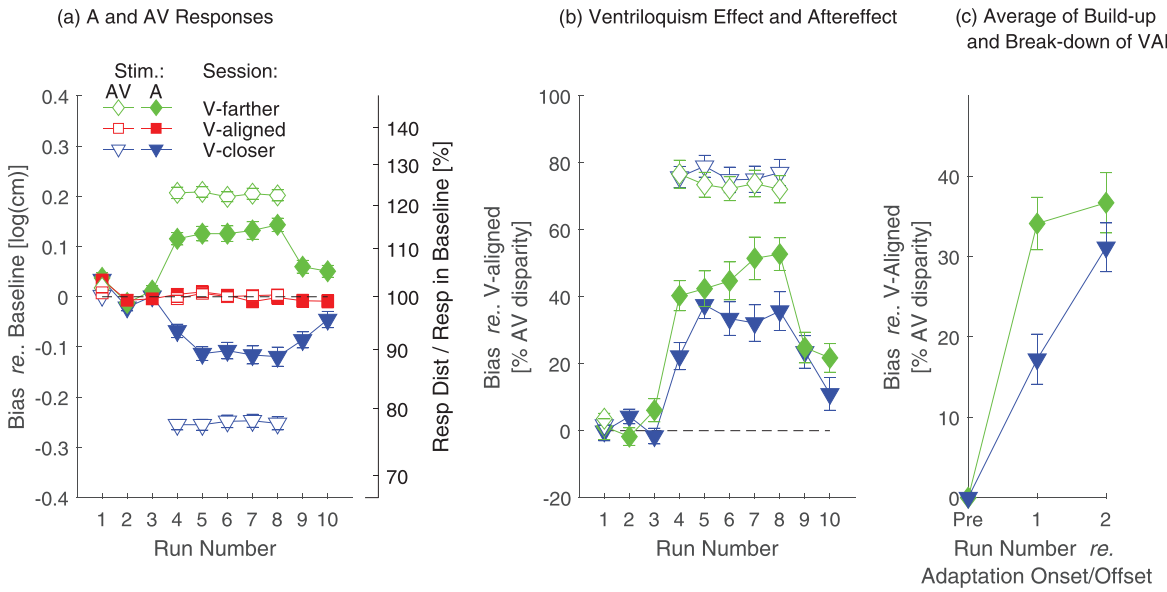

FIG. 4. (Color online) Temporal profile of ventriloquism adaptation. (A) Response biases as a function of run number within a session for the three different conditions, averaged across target distance and referenced to the pre-adaptation baselines (runs 1–3). (B) VE (open symbols) and VAE (filled symbols) as a function run number, computed from data in (A) by referencing the V-closer and V-farther data to the V-aligned data and scaling them by the physical disparity between the stimuli. (C) Dynamics of post-onset/offset VAE adaptation shown as an average of the build-up (runs 4 and 5) and inverse of break-down (runs 9 and 10) of VAE from (B), referenced to the final runs prior to the onset (runs 1–3) and offset (run 8) of the AV stimuli.

by the physical disparity of the visual and auditory components (Secs. III C and III D). Figure 4(C) aims at estimating the average dynamics of the VAE in the V-closer vs V-farther condition by showing the average of the VAE build-up [runs 4–5 referenced to the average of preadaptation runs 1–3 from Fig. 4(B)] and the inverse of the VAE break-down [runs 9–10 referenced to run 8 from Fig. 4(B)].

Figure 4(A) shows that V-aligned data, averaged across target locations, were stable near 0 log(cm) throughout the sessions (red open and filled diamonds are always near 0). The VE was fast and stable [open green diamonds and blue triangles show constant values throughout the adaptation runs 4–8 in Fig. 4(A)], while the VAE had a clear build-up and decay after the onset and offset of adaptation (full green diamonds and blue triangles grow gradually in runs 4–8 and decay gradually in runs 9–10). When expressed as a proportion of physical disparity [Fig. 4(B)], the VE magnitudes were approximately equal for the V-closer and V-farther conditions, at 70%–80% of the V-component offset (open symbols). The VAE expressed as a proportion of physical disparity was slower than the VE and had more complex dynamics. Confirming the results described in Sec. III C, VAE was stronger for the V-farther than V-closer conditions during both the adaptation and the post-adaptation runs [green filled diamonds are mostly above blue triangles for runs 4–10 in Fig. 4(B)]. Importantly, results also showed a difference in the rate of adaptation. The V-farther data [filled green diamonds in Fig. 4(B)] were approximately 40% of AV disparity by the first adaptation run (run 4), while the V-closer data [filled blue triangles in Fig. 4(B)] were only 22% of the AV disparity within the first adaptation run. Also, the V-farther data continued to grow, reaching more than 50% by run 8, while the V-closer data were approximately constant at around 35% in runs 5–10. A similar pattern was observed after the offset of the adaptation. V-farther data dropped much faster than the V-closer data (compare runs 9 vs 8 for the filled green diamonds and blue triangles). To better visualize this difference in adaptation rate, Fig. 4(C) plots the average of the build-up and the inverse of the break-down data from the two runs post-onset and post-offset of adaptation (relative to the preceding runs). The panel clearly shows much faster growth in the V-farther adaptation in the first run after its on/offset, while the V-closer adaptation grew more gradually.

Statistical analysis confirmed these results. First, an ANOVA performed on the VE data from Fig. 4(B) with the factors of run (4–8) and condition (V-closer vs V-farther) found no main effect or interaction of the condition and run factors. A similar ANOVA performed on the VAE data from Fig. 4(B) only found a significant main effect of condition [$F(4,584) = 4$; $p < 0.01$]. Finally, an ANOVA performed on data from Fig. 4(C) found a significant main effect of run [$F(1,73) = 16.94$; $p < 0.01$], a significant main effect of condition [$F(1,73) = 11.51$; $p < 0.01$], and a significant interaction of the two factors [$F(1,73) = 10.22$; $p < 0.01$].

Overall, the current VE data show that the weight used for each modality in auditory-visual distance integration is approximately constant even across multiple experimental runs. On the other hand, the VAE is slow because it entails neural adaptation of the perceptual map of auditory space guided by the visual signals. Moreover, the current results show that the neural adaptation is faster when the visual signals guide perception of auditory targets as farther away compared to when perception is guided to shift perception of auditory targets to be nearer.

https://doi.org/10.1121/10.0007066

IV. DISCUSSION

The current study examined ventriloquism effect (VA) and aftereffect (VAE) in the distance dimension for auditory stimuli presented over a range of distances from 0.7 to 2.03 m in real reverberant environment in front of the listener. It found, for a fixed 30% relative shift of the visual component of the AV stimuli, that the induced ventriloquism effect is constant at approximately 72% of the V-displacement, with no systematic variation as a function of stimulus distance or the direction of the induced shift. Also, this VE strength was constant over time. On the other hand, the ventriloquism aftereffect induced by these stimuli, while still independent of target distance, was stimulus-direction dependent. The VAE was stronger when the V-component of AV stimuli was placed farther than the auditory component (reaching 44% of the V-displacement) than when it was placed closer (31%). Also, the VAE in the V-farther condition was faster than in the V-closer condition.

The observed differences between VE and VAE confirm that these effects are caused by two different processes, comparable with the VE and VAE in the horizontal dimension (Bosen *et al.*, 2017; Bruns *et al.*, 2011; Kopčo *et al.*, 2009). Specifically, VE is likely a consequence of an immediate cross-modal integration, while VAE is a result of adaptation of some auditory spatial representation by the visual signals. For VE, the observed independence of the effect of the direction of shift is not consistent with the "proximity image effect" (Gardner, 1968) which showed that, in anechoic room, the ventriloquism effect is stronger in the V-closer direction. A possible explanation of this difference is that the current experiment was performed in a real reverberant room and the subjects were provided with a level-independent auditory distance cue, the direct-to-reverberant energy ratio, whereas in the Gardner study no such cue was available, and participants likely relied on guessing.

The observed constant 72% weight given to the V-component in the current study is much smaller than the weight typically observed in horizontal studies, in which this weight is commonly more than 90% (Bruns *et al.*, 2011; Frissen *et al.*, 2012; Kopčo *et al.*, 2009; Kopčo *et al.*, 2019; Lewald, 2002; Recanzone, 1998, 2009; Stekelenburg *et al.*, 2004; Vroomen *et al.*, 2001). Assuming that auditory and visual components were combined optimally (Alais and Burr, 2004) in the current study, this result suggests that auditory distance acuity is much more comparable to visual distance acuity than found in the horizontal dimension. However, more studies are needed to determine whether the observed 72% V-component weight generalizes to other AV disparities, other reference distances, and other types of stimuli (e.g., familiar three-dimensional objects might provide more accurate visual distance information than the single LEDs used here, while familiar sounds like speech stimuli might provide more auditory distance information).

Even though ventriloquism effect observed in this study was equally strong in the V-farther and V-closer conditions, the ventriloquism aftereffect evoked by it was stronger in the V-farther condition than in the V-closer condition. This asymmetry in the effect was not expected and it is not immediately clear why it occurred. It might be related to the "adjacency effect" reported in auditory distance perception by Min and Mershon (2005), which observed that manipulating the perceived distance of a nearby sound by a simultaneously presented distal light (i.e., ventriloquism effect) caused an adjacent reference sound to be perceived as farther away (i.e., ventriloquism aftereffect), whereas the reverse setup did not cause an adjacent reference sound to be perceived as closer. The current results suggest that this asymmetry may be a general observation in the adaptation of auditory distance percepts by vision and that it also applies to the speed with which the aftereffect builds up. Alternatively, this effect might relate to the fact that the A baseline responses were overestimated even in the pre-adaptation in the current study, which means that the perceived A-only locations were much closer to the V-farther visual adaptors than V-closer visual adaptors, even when they were interleaved with V-aligned AV stimuli. Finally, the constant 30% disparity used in this study corresponds to a larger disparity in logarithmic units for the V-closer condition (−0.36) than for the V-farther condition (0.26) [Fig. 1(B)]. In the analysis performed here, the scaling by the actual applied separation corrected for this difference. However, it is possible that the strength of AV binding was stronger in the V-farther condition, as the separation in log units was smaller, which might have resulted in the observed stronger adaptation. Future studies will need to independently control the physical and perceptual disparities in the auditory and visual components of the auditory-visual and auditory stimuli in order to disentangle these alternatives.

The VAE data of the current study show, in both V-closer and V-farther conditions, that participants adapt to a constant-ratio shift in visual-component distance over a range of distances by a shift in response that is, again, approximately constant over the examined distance range. Previous studies in the horizontal plane suggested that adaptation in response to a non-linear transformation of space often leads to a linear response pattern that approximates the non-linear transformation (Shinn-Cunningham *et al.*, 1998b). Similarly, the current results are consistent with the interpretation that the neural structure undergoing adaptation in distance ventriloquism also adapts linearly in the logarithmic units in which distance is represented, a result that is also consistent with the variance in responses to auditory stimuli which is approximately constant across distances (Anderson and Zahorik, 2014; Kopčo *et al.*, 2012). However, this interpretation does not explain why there is the observed closer-vs-farther asymmetry in the strength of the aftereffect. Also, the fixed ratio scale was the only scale examined here. I.e., we did not investigate whether the adaptation was equally good for a linearly constant shift in the V-component and nor did we examine how accurately a linear model would fit the current data. Thus, it is also possible that the representation/adaptation has other forms than

logarithmic (e.g., linear), or that it is flexible enough to be capable of adapting to several various transformations evoked by visual stimuli, as observed in the horizontal domain, where ventriloquism adaptation can be obtained both by shifting the V-component by a constant amount or by varying the gain of the V-component displacement; see Zwiers *et al.* (2003). Finally, it is even possible that the adaptation is robust to varying size of the AV displacement from trial to trial, as recently shown for horizontal ventriloquism (Bruns *et al.*, 2020).

In terms of the temporal profile, the large immediate VAE and smaller persistent VAE observed here are generally consistent with the observations of multiple distinct cross-modal adaptation mechanisms reported in the horizontal VAE studies (Bosen *et al.*, 2018; Watson *et al.*, 2019). However, a single adaptation mechanism cannot be ruled out for the current results given that the current study was not specifically designed to examine this question. On the other hand, the closer-vs-farther asymmetry in the VAE observed here differs from horizontal ventriloquism in which no equivalent left-right asymmetries are observed. Future studies will need to examine whether the observed asymmetry in the adaptation rates is directly related to the observed difference in the strength of the induced VAE, or whether these two phenomena are separable.

Our experimental setup consisted of real auditory and visual sources which assured that the stimuli were well externalized and corresponded to true physical distances. On the other hand, it also poses some limitations. The main ones are that (1) the range of A and V stimulus locations was limited by the room size, (2) the sound received from individual loudspeakers varied slightly as loudspeakers cast an acoustic shadow affecting the more distant speakers when placed behind each other especially in the high frequency region which might have caused coloration of the sound [see Fig. 1(E)]. Although coloration might have shifted the response biases in the A baseline, such bias would be accounted for by the relative comparisons and would be minimized in the effects of vision on auditory distance, (3) the visual stimuli were linearly spaced by 4.76 cm and offset by about 5 cm from the acoustic stimuli, (4) the responses were collected through the same LED array as the visual stimuli. However, (1) the room was of a common size and shape, in which listeners are expected to naturally judge distances in everyday situations, (2) the partial shadowing of sources is a common occurrence in normal listening situations and it is mainly at high frequencies while low-frequencies dominate for distance localization of broadband stimuli (Kopčo and Shinn-Cunningham, 2011), (3) auditory-visual synchrony was expected to provide a strong binding cue, and (4) using visually guided response is common in ventriloquism literature (Razavi *et al.*, 2007). Some of these limitations could be mitigated by conducting future experiments in virtual reality (VR) (Hendrikse *et al.*, 2018; Postma and Katz, 2017; Seeber *et al.*, 2010). On the other hand, our study could also have implications to auditory-visual VR experience since the perceptual effects are likely to generalize to these environments as well.

Finally, an unexpected result of the current study is that A-only responses were slightly overestimated in the pread-aptation baseline measurements and in the V-aligned condition, even though multisensory enhancement would be expected to occur, reducing the biases in the A-only stimuli (Anderson and Zahorik, 2014; Bruns *et al.*, 2020). It is difficult to identify the cause of this disparity, as no A-only initial runs were performed to serve as a baseline. Thus, it is for example possible that the enhancement has actually occurred and that the biases would have been much larger in the initial run if it was performed without any interleaved V-aligned AV stimuli, as supported by the auditory-only follow-up measurement described in Appendix A. However, even if that was the case, it still is not clear why the bias would not continue to reduce even during the five adaptation runs in the V-aligned condition. Future studies are needed to explore this effect.

V. CONCLUSION

The current results show that VE and VAE in the distance dimension can be robustly induced in real reverberant environments and that they are approximately independent of target distance on a logarithmic scale. They also show that, while VE is immediate and independent of the direction of induced shift, VAE is more complex, showing stronger and faster effects in the V-farther direction.

ACKNOWLEDGMENTS

This work was supported by VEGA 1/0355/20 and UPJŠ VVGS-2020-1514 to N.K., International Visegrad Fund 51300869 to L.H., and by BCS-1057625 to A.S.

APPENDIX A

The current study did not aim to address auditory distance perception *per se*, i.e., without considering the effect of visual stimuli. However, it is important to determine the extent to which the current experimental methods are equivalent to those used in the available auditory distance literature. To this end, this Appendix first reports the results of an auditory-only follow-up experiment performed using the same setup as in the main study. Then, it compares the current baseline results to those reported in several previous studies.

The main experiment always included auditory-visual stimuli, starting and ending with the V-aligned condition in which A-only and AV stimuli were interleaved. This included a run prior to A-only baseline measurement performed in runs 2 and 3, to improve the accuracy of participants responses, as well as to avoid any auditory-visual carry-over effects. As a follow-up experiment was performed with the same methodology as used in the main experiment, with the exception that all 11 runs of two sessions were always A-only conditions (i.e., there was no visual component in the stimuli). This analysis, averaged

https://doi.org/10.1121/10.0007066

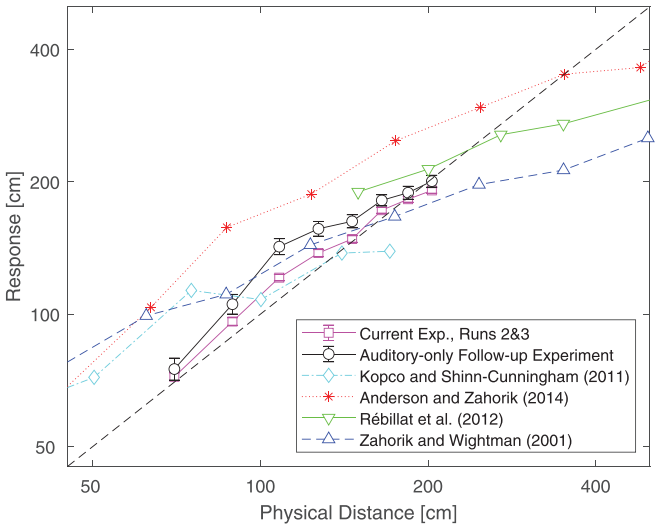

FIG. 5. (Color online) Results of a follow-up auditory-only experiment compared to the auditory-only baseline from the main study and to several results from the literature. For the current experiments, across-subject geometric mean of the response distance (±SEM) as a function of actual target distance is reported.

data across two sessions and across 22 subjects who participated in the follow-up.

Figure 5 presents the results of this follow-up along with the results of the baseline A-only measurement in the main study and with several comparable auditory distance experiments from the literature. The geometric mean of response distance from the auditory-only follow-up is shown by circles. Data from the A-only runs 2 and 3 from the main experiment, averaged across groups and sessions are shown by squares. Data were for relevant egocentric distances for frontal broadband targets from Kopčo and Shinn-Cunningham (2011), Fig. 2(A) broadband condition ($n = 6$), shown by diamonds, Anderson and Zahorik (2014), Fig. 2(A) ($n = 57$), shown by asterisks, Rébillat *et al.* (2012), Fig. 8(B), position 1 ($n = 40$), shown by down-pointing triangles, and Zahorik and Wightman (2001), Fig. 5 ($n = 5$), shown by up-pointing triangles.

The data of the follow-up experiment are slightly more biased towards larger distances than the data of the main experiment, which confirms the assumption that the V-aligned condition in the first run of the main experiment made the responses better aligned with the true loudspeaker positions. For the distances 100–200 cm, the data of the current experiments (both main and the follow-up) are very close to the data of Zahorik and Wightman (2001) as well as to (Kopčo and Shinn-Cunningham, 2011). While Anderson and Zahorik (2014) and Rébillat *et al.* (2012) show overestimation. For the distance below 100 cm, the current data are less biased than Zahorik and Wightman (2001) or Kopčo and Shinn-Cunningham (2011), suggesting that the listeners were less biased in the current study. The reason for this improvement might be that the acoustic high-frequency shadowing enhanced the distance-dependent spectral-coloration cue, which counteracted the overestimation bias observed in the previous studies. Or, it is possible that the listeners were able to judge those sources much more accurately in the current study's real environment than in the virtual environment employed previously.

Previous studies modeled perception of auditory distance using power law function ($r' = kr^a$, r' – estimate of perceived distance; r, physical sound source distance; k, a, parameters) (Anderson and Zahorik, 2014; Zahorik *et al.*, 2005). As an additional effort to establish that the baseline performance here is comparable with corresponding conditions of previous studies, the model was also fitted to the current A-only and AV responses in the main-experiment baseline runs. For the A-only data, the fitted values of the parameters were $k = 1.04$ and $a = 0.82$, very near the ideal values of 1 (all values are across-subject means). For the responses to AV stimuli, the fitted parameters were $k = 1.15$, $a = 0.93$. Both these fits are in the range of typical values found across literature (Zahorik *et al.*, 2005). A study which used an array of loudspeakers found values of $k = 0.78$, $a = 0.9$ (in condition with all speakers visible) and $k = 0.92$, $a = 0.66$ (in condition with the participants blindfolded) (Zahorik, 2001). The visible-speaker condition could be directly compared with our A-only condition, and the observed values of k and a are comparable. In another study that used head-phone presentation and video screen presentation the AV condition was fitted with values of $k = 0.96$,

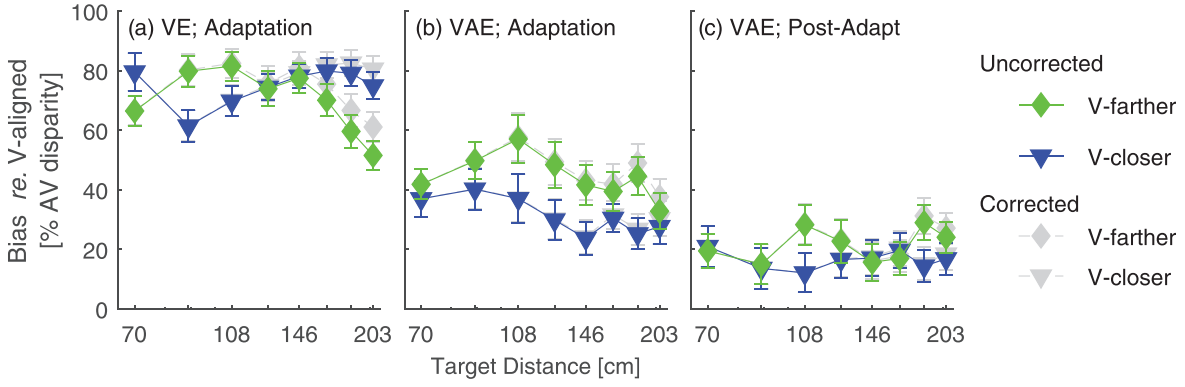

FIG. 6. (Color online) Ventriloquism effect and aftereffect with and without the correction for the visual baseline. The corrected data (in gray with dashed lines) are re-plotted from Figs. 3(G)–3(I). The uncorrected data are plotted identically to the data in Fig. 3 but without the correction for the visual baseline. For the color version of the figure, please refer to the online version.

TABLE I. Summary statistical analysis of the VE, VAE, and persistent VAE for the uncorrected data shown in Fig. 6. *$p < 0.05$, ***$p < 0.001$ (significance levels modified by Geisser-Greenhouse Epsilons).

	VE [Fig. 6(A)]	VAE [Fig. 6(B)]	Persistent VAE [Fig. 6(C)]
Target Location F(71 022)	2.7*	2.58*	0.57
condition F(1,146)	0.89	4.67*	0.63
Target Location x condition F(71 022)	6.92***	0.8	1.28

$a = 0.93$ (Anderson and Zahorik, 2014). These fitted values are comparable with our AV condition. On the other hand, a study which used VR and a method of triangulation for auditory-visual distance estimations found more compressed responses ($a = 0.41$ or $a = 0.29$, $k = 1.62$, or $k = 2.13$) (Rébillat *et al.*, 2012). These values differ more considerably from the current results. Since our results obtained in real environment approach the ideal values of 1, it is likely that the inferior performance in this study can be ascribed to the use of virtual environment, to different distance ranges, or to other differences in the experimental design.

Overall, the comparison provided here shows that the current baseline auditory-only performance is comparable to the previous studies. The responses are more accurate for the observed range than in the previous studies, which could relate to the availability of the direct sound level cue, loudspeakers in a real room (as opposed to virtual environments), or the additional spectral cues caused by shadowing of the loudspeakers. Importantly for the current audiovisual study, the fact that the auditory distance percepts are less biased here, i.e., that the auditory-only and visual-only percepts are closer to each other in the V-aligned condition, might mean that the audio-visual binding was stronger here than would occur in virtual environment.

APPENDIX B

Performance in the baseline V-only condition was corrected for the perceptual biases relating to the setup used in the present study. The visual stimuli in baseline condition were not compensated for any artifacts that may have resulted in the specific placement of our stimuli and were not tested against perceptually neutral reference. Although the resulting correction was small in the magnitude, it slightly influenced the magnitudes of reported ventriloquism effect and aftereffect. The effect of the correction is shown in Fig. 6, which shows both corrected and uncorrected results. Overall, the effect is small, mainly affecting the V-farther condition.

Table I shows the results of ANOVA for all three effects with uncorrected data. The main difference between the uncorrected and corrected data in terms of the statistical analysis is that the ANOVA for the uncorrected data showed significant effects of target location for the VE and VAE in addition to the effects reported with the corrected data. In terms of VE, the effects are mainly driven by the interaction and, as with the corrected data, there is no clear trend in terms of the magnitude of VE growing/decreasing with increasing egocentric distance. Therefore, it is likely that the significant effect of target location is driven by the measurement artifacts eliminated by the correction. In terms of VAE, the change of the data were about five percentual points for the two farthest target targets in the V-farther condition. In other target locations it was even smaller. This is relatively small change when averaged across all positions, however, due to a high number of participants even small change could change the significance levels of the statistical test because one number is added to the data of all participants. Again, the change due to the correction was relatively small but it seems to correct for the effects that most likely relate to the setup of the current study.

Alais, D., and Burr, D. (**2004**). "The ventriloquist effect results from near-optimal bimodal integration," Curr. Biol. **14**, 257–262.

Anderson, P. W., and Zahorik, P. (**2014**). "Auditory/visual distance estimation: Accuracy and variability," Front. Psychol. **5**, 1–11.

Bedford, F. L. (**1993**). "Perceptual and cognitive spatial learning," J. Exp. Psychol. Hum. Percept. Perform. **19**, 517–530.

Bertelson, P., Frissen, I., Vroomen, J., and de Gelder, B. (**2006**). "The aftereffects of ventriloquism: Patterns of spatial generalization," Percept. Psychophys. **68**, 428–436.

Bertelson, P., Vroomen, J., de Gelder, B., and Driver, J. (**2000**). "The ventriloquist effect does not depend on the direction of deliberate visual attention," Percept. Psychophys. **62**, 321–332.

Berzborn, M., Bomhardt, R., Klein, J., Richter, J.-G., and Vorländer, M. (**2017**). "The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing," in Proceedings of the DAGA 2017 - 43th Annual German Congress on Acoustics, 6–9 March 2017, Kiel, Germany.

Bosen, A. K., Fleming, J. T., Allen, P. D., O'Neill, W. E., and Paige, G. D. (**2017**). "Accumulation and decay of visual capture and the ventriloquism aftereffect caused by brief audio-visual disparities," Exp. Brain Res. **235**, 585–595.

Bosen, A. K., Fleming, J. T., Allen, P. D., O'Neill, W. E., and Paige, G. D. (**2018**). "Multiple time scales of the ventriloquism aftereffect," PLoS One **13**, e0200930.

Brungart, D. S., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources I: Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

Bruns, P., Dinse, H. R., and Röder, B. (**2020**). "Differential effects of the temporal and spatial distribution of audiovisual stimuli on cross-modal spatial recalibration," Eur. J. Neurosci. **52**, 3763–3775.

Bruns, P., Liebnau, R., and Röder, B. (**2020**). "Differential effects of the temporal and spatial distribution of audiovisual stimuli on cross-modal spatial recalibration," Eur. J. Neurosci. **52**, 3763–3775.

Calcagno, E. R., Abregú, E. L., Eguía, M. C., and Vergara, R. (**2012**). "The role of vision in auditory distance perception," Perception **41**, 175–192.

Cubick, J., Santurette, S., Laugesen, S., and Dau, T. (**2015**). "The influence of visual cues on auditory distance perception," DAGA 2015 - 41. Annu. Ger. Congr. Acoust. 16–19 March 2015, Nürnberg, Germany, 1220–1223.

Frissen, I., Vroomen, J., and de Gelder, B. (**2012**). "The aftereffects of ventriloquism: The time course of the visual recalibration of auditory localization," Seeing Perceiv. **25**, 1–14.

Gardner, M. B. (**1968**). "Proximity image effect in sound localization," J. Acoust. Soc. Am. **43**, 163–163.

Hartmann, W. M. (**1983**). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.
Hendrikse, M. M. E., Llorach, G., Grimm, G., and Hohmann, V. (**2018**). "Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters," Speech Commun. **101**, 70–84.
Herron, T. (**2005**). "C Language exploratory analysis of variance with enhancements," http://www.ebire.org/hcnlab/software/cleave.html Last Viewed: 1 November 2021.
Jack, C. E., and Thurlow, W. R. (**1973**). "Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect," Percept. Mot. Skills **37**, 967–979.
Kopčo, N., Huang, S., Belliveau, J. W., Raij, T., Tengshe, C., and Ahveninen, J. (**2012**). "Neuronal representations of distance in human auditory cortex," Proc. Natl. Acad. Sci. **109**, 11019–11024.
Kopčo, N., Lin, I.-F., Shinn-Cunningham, B. G., and Groh, J. M. (**2009**). "Reference frame of the ventriloquism aftereffect," J. Neurosci. **29**, 13809–13814.
Kopčo, N., Lokša, P., Lin, I., Groh, J., and Shinn-Cunningham, B. (**2019**). "Hemisphere-specific properties of the ventriloquism aftereffect," J. Acoust. Soc. Am. **146**, EL177–EL183.
Kopčo, N., and Shinn-Cunningham, B. G. (**2011**). "Effect of stimulus spectrum on distance perception for nearby sources," J. Acoust. Soc. Am. **130**, 1530–1541.
Lewald, J. (**2002**). "Rapid Adaptation to Auditory-Visual Spatial Disparity," Learn. Mem. **9**, 268–278.
Mendonça, C., Mandelli, P., and Pulkki, V. (**2016**). "Modeling the perception of audiovisual distance: Bayesian causal inference and other models," PLoS One **11**, e0165391.
Mershon, D. H., Desaulniers, D. H., and Amerson, J. (**1980**). "Visual capture in auditory distance perception: Proximity image effect reconsidered," J. Aud. Res. **20**, 129–136.
Mershon, D. H., Desaulniers, D. H., Kiefer, S. A., Amerson, J., and Mills, J. T. (**1981**). "Perceived loudness and visually-determined auditory distance," Perception **10**, 531–543.
Min, Y. K., and Mershon, D. H. (**2005**). "An adjacency effect in auditory distance perception," Acta Acust. united Ac. **91**, 480–489.
Postma, B. N. J., and Katz, B. F. G. (**2017**). "The influence of visual distance on the room-acoustic experience of auralizations," J. Acoust. Soc. Am. **142**, 3035–3046.
Razavi, B., O'Neill, W. E., and Paige, G. D. (**2007**). "Auditory spatial perception dynamically realigns with changing eye position," J. Neurosci. **27**, 10249–10258.
Rébillat, M., Boutillon, X., Corteel, É., and Katz, B. F. G. (**2012**). "Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments," ACM Trans. Appl. Percept. **9**, 1–17.
Recanzone, G. H. (**1998**). "Rapidly induced auditory plasticity: The ventriloquism aftereffect," Proc. Natl. Acad. Sci. U.S.A. **95**, 869–875.
Recanzone, G. H. (**2009**). "Interactions of auditory and visual stimuli in space and time," Hear. Res. **258**, 89–99.
Rife, D. D., and Vanderkooy, J. (**1989**). "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–444.
Seeber, B. U., Kerber, S., and Hafter, E. R. (**2010**). "A system to simulate and reproduce audio–visual environments for spatial hearing research," Hear. Res. **260**, 1–10.
Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998a**). "Adapting to supernormal auditory localization cues I: Bias and resolution," J. Acoust. Soc. Am. **103**, 3656–3666.
Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998b**). "Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response," J. Acoust. Soc. Am. **103**, 3667–3676.
Shinn-Cunningham, B. G., Kopčo, N., and Martin, T. J. (**2005**). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. **117**, 3100–3115.
Stekelenburg, J. J., Vroomen, J., and de Gelder, B. (**2004**). "Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity," Neurosci. Lett. **357**, 163–166.
Tresilian, J. R., Mon-Williams, M., and Kelly, B. M. (**1999**). "Increasing confidence in vergence as a cue to distance," Proc. R. Soc. London. Ser. B Biol. Sci. **266**, 39–44.
Vanderkooy, J. (**1994**). "Aspects of MLS measuring systems," J. Audio Eng. Soc. **42**, 219–231.
Voss, P. (**2016**). "Auditory spatial perception without vision," Front. Psychol. **7**, 1–7.
Vroomen, J., Bertelson, P., and de Gelder, B. (**2001**). "The ventriloquist effect does not depend on the direction of automatic visual attention," Percept. Psychophys. **63**, 651–659.
Watson, D. M., Akeroyd, M. A., Roach, N. W., and Webb, B. S. (**2019**). "Distinct mechanisms govern recalibration to audio-visual discrepancies in remote and recent history," Sci. Rep. **9**, 8513.
Wozny, D. R., and Shams, L. (**2011**). "Recalibration of auditory space following milliseconds of cross-modal discrepancy," J. Neurosci. **31**, 4607–4612.
Zahorik, P. (**2001**). "Estimating sound source distance with and without vision," Optom. Vis. Sci. **78**, 270–275.
Zahorik, P. (**2003**). "Auditory and visual distance perception: The proximity image effect revisited," J. Acoust. Soc. Am. **113**, 2270–2270.
Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta Acust. united Ac. **91**, 409–420.
Zahorik, P., and Wightman, F. L. (**2001**). "Loudness constancy with varying sound source distance," Nat. Neurosci. **4**, 78–83.
Zwiers, M. P., Van Opstal, A. J., and Paige, G. D. (**2003**). "Plasticity in human sound localization induced by compressed spatial vision," Nat Neurosci. **6**, 175–181.