Frequency transfer of the ventriloquism aftereffect

Rachel Ege[1], A. John van Opstal[1], Marc M. van Wanrooij[1*]

[1]Radboud University, Donders Institute for Brain, Cognition and Behaviour, Department of Biophysics, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

*corresponding author: m.vanwanrooij@donders.ru.nl

Abstract

The ventriloquism aftereffect (VAE) describes the persistent shift of perceived sound location after having been adapted to a ventriloquism condition, in which the sound was repeatedly paired with a displaced visual stimulus. In the latter case, participants consistently mislocalize the sound in the direction of the visual stimulus (ventriloquism effect, VE). Previous studies provide conflicting reports regarding the strength of the VAE, ranging from 0 to nearly 100%. Moreover, there is controversy about its generalization to different sounds than the one inducing the VE, ranging from no transfer at all, to full transfer across different sound spectra. Here, we imposed the VE for three different sounds: a low-frequency and a high-frequency narrow-band noise, and a broadband Gaussian white noise (GWN). In the adaptation phase, listeners generated fast goal-directed head movements to localize the sound, presented across a 70 deg range in the horizontal plane, while ignoring a visual distracter that was consistently displaced 10 deg to the right of the sound. In the post-adaptation phase, participants localized narrow-band sounds with center frequencies from 0.5 to 8 kHz, as well as GWN, without the visual distracter. Our results show that the VAE amounted to approximately 40% of the VE and generalized well across the entire

frequency domain. We also found that the strength of the VAE correlated with the pre-adaptation sound-localization performance. We compare our results with previous reports and discuss different hypotheses regarding optimal audio-visual cue integration.

Significance statement

Where we believe sounds to emanate from is heavily influenced by what we see. We experience this illusion - the ventriloquism effect - regularly. For example, when watching movies, we perceive the voices of actors to originate from the actors' lips rather than from the speakers. Long exposure to a spatial mismatch between what we hear and what we see, leads to a recalibration of our localization of sounds - the ventriloquism aftereffect. Unclear was whether exposure to an audiovisual mismatch for one sound type would also lead to adaptive changes for other sounds. The present study shows that the ventriloquism aftereffect generalizes, implying that it does not act at an early auditory stage, but at a late multisensory stage.

Introduction

The ventriloquism effect (VE) describes the well-known audio-visual illusion that the spatial percept of a sound is biased towards the location of a simultaneously presented nearby visual distracter (Jack and Thurlow, 1973; Alais and Burr, 2004). This effect can be observed, for example, when lip movements of a ventriloquist's dummy are synchronous with sounds emitted by the ventriloquist. Current ideas hold that the VE is due to a statistically optimal integration of multisensory cues within the perceptual system, in which the contributions of the sensory inputs to their respective spatial percepts are weighted by their uncertainties (Ernst and Banks, 2002; Alais and Burr, 2004; Van Opstal, 2016). Because visual stimuli often induce more precise spatial percepts than sounds, the audiovisual percept typically falls at a weighted averaged location between the unimodal perceived visual and auditory locations that is biased towards the visual cue (Alais and Burr, 2004).

The so-called ventriloquism aftereffect (VAE; Canon, 1970) manifests after prolonged, consistent, and repeated presentation of an audiovisual stimulus that induces the VE. In the VAE, the spatial percept of the sound continues to be biased towards the displaced visual cue, even without the visual stimulus (Canon, 1970; Radeau and Bertelson, 1974; Recanzone, 1998; Lewald, 2002; Frissen et al., 2003).

Although the VAE has been described in numerous studies (e.g., Canon, 1970; Jack and Thurlow, 1973; Radeau and Bertelson, 1974; Recanzone, 1998; Frissen et al., 2003, 2005, 2012; Woods and Recanzone, 2004; Kopco et al., 2009; Bruns and Röder, 2015, 2017; Bosen et al., 2017,2018; Zierul et al., 2017; Berger and Ehrsson, 2018), its underlying neural mechanisms are still unclear. For example, there is disagreement about the size of the effect, and about its generalization, or transfer, to different sounds that had not been used to imprint the VE. One potential problem in quantifying the size of the VE, and hence of the VAE, lies in the fact that the perceived location of a single unimodal auditory target is not necessarily veridical. So, despite the true (physical) spatial disparity of the visual distracter with respect to the sound's location, the *perceived* spatial disparity, and hence its potential to induce a VE, may be quite different, and may vary systematically with the sound's location and its spectral content. Second, it is unclear whether the VAE is due to audiovisual interactions at a central multisensory level where all sensory stimuli have been brought into a common spatial reference frame (e.g., Kopco *et al.*, 2009), or to neural interactions within the early auditory pathways, which may still be tonotopically, rather than spatially, organized (e.g., Recanzone, 1998; Bruns and Röder, 2015, 2017).

Here, we specifically addressed these issues by investigating how the VE and VAE depend on the spectral characteristics of the co-occurring sound. We considered different potential mechanisms (schematically illustrated in Fig. 1) that could induce a VAE: (i) the tonotopic hypothesis, (ii) the localization cue hypothesis, and (iii) the spatial hypothesis. As in most studies of the VE and VAE, we here considered the localization of a narrow-band sound in the horizontal plane (the sound's

azimuth angle), which was paired with a simple synchronous visual distracter (e.g., an LED) that was displaced at a fixed distance from the auditory target.

The tonotopic hypothesis (Fig. 1A) then assumes that the audiovisual (AV) interactions between sound and visual stimulus take place at a level where the auditory input is still tonotopically represented. The VAE is then expected to only occur for sound frequencies that lie within the critical spectral band around the tone (e.g., Schooneveldt and Moore, 1989; Moore, 2003), and that there will be no transfer to other frequencies (Recanzone, 1998; Bruns and Röder, 2015, 2017).

According to the localization-cue hypothesis (Fig. 1B), the AV interactions take place at a stage in the auditory system where the binaural sound-localization cues have already been extracted. For sound frequencies up to about 1.5 kHz, the dominant localization cue is the interaural time difference (ITD), whereas for frequencies above about 3 kHz sound localization is mainly based on interaural level differences (ILD). According to this model, the VAE will transfer among all frequencies that share the same localization cue, but not to frequencies of the other localization cue. In other words, the VAE has partial frequency transfer.

The spatial hypothesis holds that AV interactions occur at a level where all sensory stimuli are represented within a common spatial representation, e.g. an eye-centered motor map (Fig. 1C; Corneil et al., 2002; Kopco et al., 2009; Van Opstal, 2016). In that case, the VAE will be expected to transfer across all frequencies (Frissen et al., 2003, 2005). Support for this hypothesis was also provided by neurobiologically plausible neural-network simulations (Magosso et al., 2012, 2013).

So far, studies have been limited in assessing frequency transfer in the VAE, as typically only a few frequencies were tested (e.g., 750 Hz and 3 kHz; Radeau and Bertelson, 1974; Recanzone, 1998; Frissen et al., 2003, 2012), which does not allow one to dissociate the cue-hypothesis and spatial hypothesis. Thus, to obtain a better understanding of the VAE, and its potential dependence on sound frequency, we extended these previous studies, by testing the VAE across a larger range of frequencies and by accounting for non-veridical spatial percepts of the sounds.

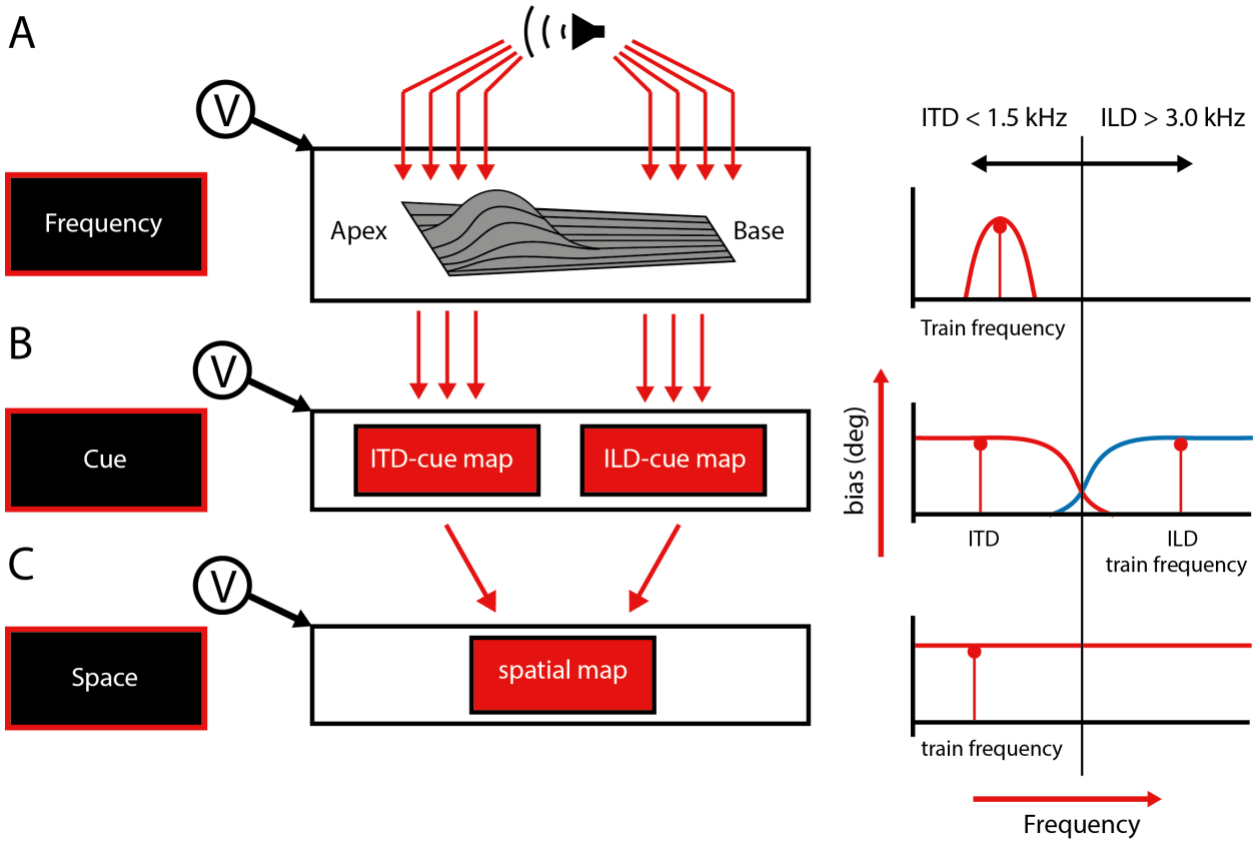

Fig. 1. Ventriloquism Aftereffect Models. *The VE is induced by a pure tone (red vertical line with red dot on the right). The predictions of three different models for VAE transfer are indicated by the red and blue curves. The thin vertical line represents the border between the ITD and ILD domains (around 2 kHz).* ***A)*** *Visual adaptation takes place at a tonotopic stage, before binaural integration of the spatial cues (left); VAE is only found for frequencies within a critical band around the adapting tone (right).* ***B)*** *Visual co-exposure leads to adaptation within the ITD- or ILD-localization cues; the VAE will be restricted to frequencies belonging to that cue (right, center).* ***C)*** *Generalization across the entire frequency domain occurs if visual exposure leads to adaptation at a spatial map (right, bottom).*

Materials and Methods

Participants

Nine human participants were recruited at a location which will be identified if the article is published. The data of three participants were excluded for further analysis. One participant demonstrated poor sound localization for low-frequency sounds, and two participants did not demonstrate a ventriloquism effect during the exposure block. The data of six participants remained (three females; age 21-28 yrs.; mean = 25.7 years). All participants had normal or corrected-to-normal vision and normal hearing (verified by means of a standard sound localization experiment including low-pass, high-pass and Gaussian white noise sounds in the frontal hemifield, and an audiogram). One participant is the first author of this study; the other participants were kept naive about the purpose of the study. Experiments were conducted after obtaining informed consent from the participant. The experimental protocol was approved by the local ethics committee of the University (protocol nr. will be identified if the article is published), and all applied procedures were in line with the relevant guidelines and regulations.

Apparatus

During the experiment, the participant sat comfortably in a chair in the center of a completely dark, sound-attenuated room. The room had an ambient background noise level below 30 dBA. The chair was positioned at the center of a spherical frame (radius 1.5 m), on which 125 small broad-range loudspeakers (SC5.9; Visaton GmbH, Haan, Germany) were mounted. These speakers were organized in a grid by separating them from the nearest speakers by an angle of approximately 15 degrees in both azimuth and elevation according to the double-pole coordinate system (Knudsen and Konishi, 1979). Along the cardinal axes, speakers were separated by 5 deg. Head movements were recorded with the magnetic search-coil technique (Robinson, 1963). To this end, the participant wore a lightweight spectacle frame with a small coil attached to its nose bridge. Three orthogonal pairs of square coils (6 mm^2 wires, 3 x 3 m) were attached to the room's edges to generate the horizontal (80 kHz), vertical (60 kHz) and frontal (48 kHz) magnetic fields, respectively. The horizontal and vertical head-coil signals were amplified and demodulated (EM7; Remmel Labs, Katy, TX, USA), low-pass-filtered at 150 Hz (custom built, fourth-order

Butterworth), digitized (RA16GA and RA16; Tucker-Davis Technology, Alachua, FL, USA) and stored on hard disk at 6000 Hz/channel. A custom-written Matlab program, running on a PC (HP EliteDesk, California, United States) controlled data recording and storage, stimulus generation, and online visualization of the recorded data.

Stimuli

Acoustic stimuli were digitally generated by Tucker-Davis System 3 hardware (Tucker-Davis Technology, Alachua, FL, USA), consisting of two real-time processors (RP2.1, 48,828.125 Hz sampling rate), two stereo amplifiers (SA-1), four programmable attenuators (PA-5), and eight multiplexers (PM-2).

We presented eight sound types during the experiment: seven band-limited Gaussian white noise sounds with different center frequencies (cut-off frequencies, one-third octave above and below the center frequency): 500, 793, 1260, 2000, 3175, 5040, 8000 Hz, and one broadband GWN (0.5 – 20 kHz). Each sound had a duration of 75 ms, was pre-generated and stored on disk. It was presented at 55 dBA, and had 5-ms sine-squared onset, cosine-squared offset ramps. While each sound type was presented during the pre-exposure and post-exposure blocks of each session, during the exposure block solely one of three sounds (793 Hz (low frequency), 5040 Hz (high frequency), and GWN (broadband)) was presented.

Visual stimuli consisted of green LEDs (wavelength 565 nm; Kingsbright Electronic Co., LTD., Taiwan; luminance 1.4 cd/m^2) that were mounted at the center of each speaker in the setup, and served as independent visual fixation stimuli during the calibration experiment, as a central fixation light at straight ahead during the sound-localization experiments, or as a visual distracter during the ventriloquism adaptation block (stimulus duration: 75 ms).

Calibration

To determine the off-line calibration file that maps the raw coil signals (in mV) onto known target locations (in deg), participants pointed their head towards 24 LED locations in the frontal hemifield (separated by approximately 30 deg in both azimuth and elevation), using a red laser,

which was attached to the spectacle frame at a fixed distance of 40 cm in front of the eyes. A three-layer neural network, implemented in Matlab, was trained to carry out the required mapping of the raw initial and final head position signals onto the (known) LED azimuth and elevation angles with an overall accuracy of 1.0 deg, or better. The weights of the neural network were subsequently used to map all head-movement voltages of the localization experiments to degrees.

Experimental Design and Statistical Analysis

Paradigms

Each participant participated in three sessions and each session consisted of three blocks (pre-exposure, peri-exposure, and post-exposure). The pre- and post-exposure blocks were the same, except for the order of stimulus-location combinations. Here, all eight sounds were presented at eight locations (-35 to +35 deg, in 10 deg steps; Fig. 2A) in the horizontal plane. Each stimulus-location combination was presented four times in pseudo-random order (8x8x4 = 256 trials). During the exposure block one of three sounds (793 Hz [low frequency], 5040 Hz [high frequency], and GWN [broadband]) was presented at the same eight locations (-35 to +35 deg, in steps of 10 deg, Fig. 2A), while simultaneously a visual stimulus was presented ten degrees to the right (-25 to +45 deg, in steps of 10 deg, Fig. 2A). Each combination was repeated seven times (8x56 = 448 trials). In each block, participants first aligned the head-fixed laser with a central fixation light. They then pressed a button, whereupon the fixation light was extinguished (random timing 300-800 ms after button press). Either the single sound (pre- and post-exposure) or the sound-light combination (exposure) was presented 200 ms after the fixation light disappeared and participants were asked to align the head-fixed laser pointer as quickly and as accurately as possible with the sound source's origin (neglecting the visual stimulus). One session, including the calibration and short breaks in between the blocks took approximately 60 minutes. Sessions were performed on separate days.

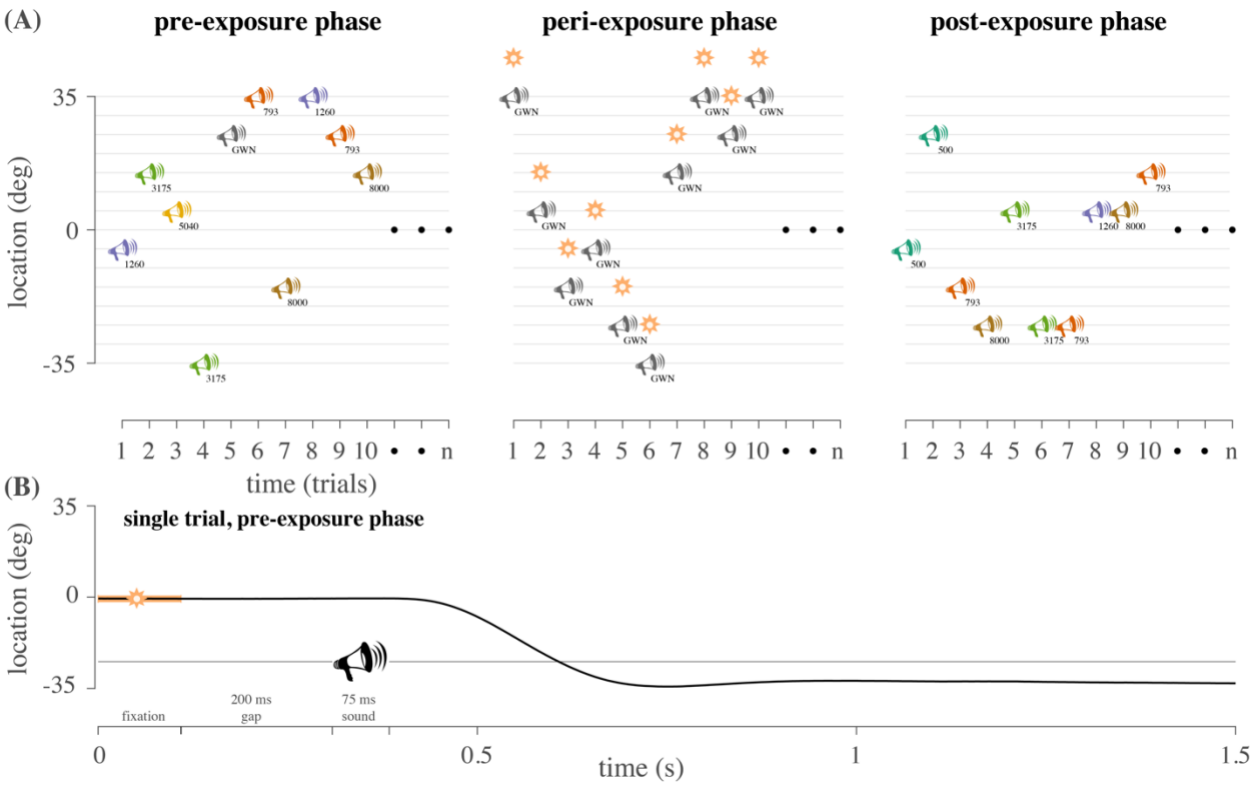

Fig. 2. Experimental paradigm. ***(A)*** *Experimental timeline. During the pre-exposure (left) and post-exposure (right) blocks, participants localized sounds (denoted by a speaker symbol) of eight different frequency bands (indicated by different colors, with center frequencies denoted by text) at eight different locations in the horizontal plane. During the exposure phase (center), a sound of a single frequency band was presented, in synchrony with a visual distractor (indicated by the yellowish sun symbol) 10 deg to the right of the sound source. Shown here are the first 10 trials of each experimental phase in a session of a single participant, in which the GWN was the target sound in the exposure phase.* ***(B)*** *Trial sequence. Each trial started with a central fixation light, which the participant fixated by aligning the head-fixed laser. After pressing a button, the fixation LED extinguished after a random period, and a peripheral sound (with visual distracter in the exposure block) was presented 200 ms later. The participant was instructed to localize the target sound, while ignoring any potential visual distractor, as quickly and as accurately as possible with a head-orienting response (bold black line).*

Veridical vs non-veridical localization

It is typically assumed in the analysis of the VAE that sound localization performance is veridical. Then, the VAE amounts to a constant bias in the direction of the visual distracter, which is quantified by the mean unsigned error (the linear offset between sound target and response). However, many sound-localization studies have shown that the perceived location of band-limited noise sources may be quite inaccurate (Musicant and Butler, 1984; Middlebrooks, 1992; Goossens and Van Opstal, 1999; Van Grootel et al., 2011). As a result, a visual distracter with a 10 deg shift to the right of an auditory target will not be perceived displaced at 10 deg to the right, but could even be perceived to the left of the sound (see Fig. 3B). Therefore, in quantifying the VAE, both the localization gain (slope) and offset (bias) of the stimulus-response relation need to be accounted for. One way in which the listener could deal with location-dependent audio-visual mismatch, is changing gain and bias during post-adaptation (Fig. 3). Here, we propose a multiple linear regression model that deals with non-veridical localization data and provides a more robust estimate of the VAE than the putative constant unsigned error.

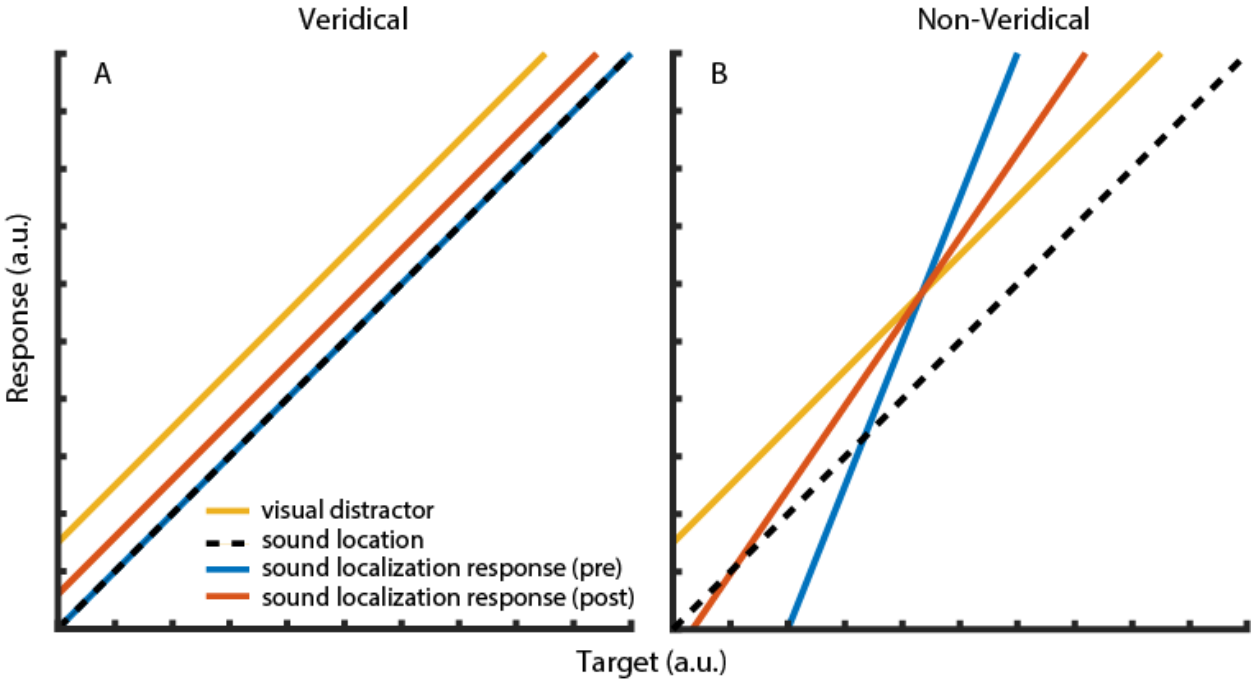

Fig. 3. Potential effects of veridical and non-veridical sound localization. ***A)*** *When sound localization is veridical (gain = 1.0, bias = 0 deg; blue line; black dashed line: sound location), participants only need to change their response bias (orange) to reduce the audiovisual mismatch with the visual distracter (yellow).* ***B)*** *If sound localization is not veridical (blue; gain ≠ 1.0, and bias ≠ 0 deg), the (perceived) audiovisual offset (distance between the yellow and blue lines) depends on the sound's location and may even change sign. Changing only the response bias would typically lead to larger errors. Adjusting the response gain and bias such that the 'perceived audiovisual' error is reduced (e.g., by a clockwise rotation of the blue line around the intersection point (orange)), would be an optimal way to deal with this problem.*

Estimating the ventriloquism effect and after effect

We described the localization responses $\hat{R}_{A,s}$ for sound *s* as a linear function of the auditory target location, $T_{A,s}$:

$$\hat{R}_{A,s} = g_s \cdot T_{A,s} \qquad \text{Eq. 1}$$

in which the dimensionless response gain, g_s, is the slope of the stimulus-response relationship for sound s. We did not include an offset in Eq. 1, which would reflect a generic response bias, as this was found to be typically negligible (close to zero deg) for all participants.

The visual-induced response bias, $\hat{B}_{VE,v}$, induced in the ventriloquism condition for the target sound *v* at location $T_{A,v}$ (here, *v* = 793 Hz, 5040 Hz, or GWN), and a synchronous visual distracter stimulus at D_V, was described as:

$$\hat{B}_{VE,v} = \mu_v \cdot \left(D_V - \hat{R}_{A,v}\right) \qquad \text{Eq. 2}$$

with $D_V - \hat{R}_{A,v}$ quantifying the perceived spatial disparity (in deg) between the visual distracter location and the perceived auditory response location ($g_v \cdot T_{A,v}$, from Eq. 1), and μ_v the magnitude of the ventriloquism effect (dimensionless), quantified as a fraction of the perceived audiovisual spatial disparity. Subsequently, the ventriloquism aftereffect for (potentially different) sound, s, was determined by

$$\hat{B}_{VAE,sv} = \lambda_{sv} \cdot \hat{R}_{VE,v} \qquad \text{Eq. 3}$$

where λ_{sv} is the relative magnitude of the ventriloquism aftereffect, quantified as a fraction of the (presumably) persistent ventriloquism effect, which was induced by sound *v* (Eqn. 2).

The expected sound-localization response, $\hat{R}_s$, in any experimental block for sound *s* is then given by the linear sum of the auditory response and the visual bias of the ventriloquism (after)effect, by combining Eqns. 1 and 3:

$$\hat{R}_s = \lambda_{sv}\mu_v \cdot \left(D_V - g_s \cdot T_{A,s}\right) + g_s \cdot T_{A,s} \qquad \text{Eq. 4}$$

We fitted the three free parameters (g_s, μ_v, λ_{sv}) of this multiple linear regression model on the data for every sound and participant separately, but pooled for the three experimental blocks (pre-adaptation, adaptation (i.e., visual exposure), and post-adaptation block). The best-fit sound localization gain, g_s, was determined from the data of all blocks, and assumed constant during the three blocks. To avoid meaningless parameter estimates, we constrained the gains for the ventriloquism effect, μ_v, and for the ventriloquism aftereffect, λ_{sv}, for the different blocks as follows:

$$\mu_v = \begin{cases} 0, & pre \\ \mu_v, & exp \\ \mu_v, & post \end{cases} \quad \text{and} \quad \lambda_{sv} = \begin{cases} 0, & pre \\ 1, & exp \\ \lambda_{sv}, & post \end{cases} \qquad \text{Eq. 5}$$

In this way, parameters that are not meaningful for a particular block were not affected by data from that block (e.g. μ_v has no meaning in the pre-exposure sound-localization block, since there was no visual stimulus; similarly, before and during adaptation, there is no aftereffect). Substituting the appropriate values into Eqs. 4 and 5 yields:

$$\hat{R}_s = \begin{cases} g_s T_{A,s} & pre, A = s \\ \mu_v \cdot \left(D_V - g_v T_{A,v}\right) + g_v T_{A,v} & exp, A = v \\ \lambda_{sv}\mu_v \cdot \left(D_V - g_s T_{A,s}\right) + g_s T_{A,s} & post, A = s \end{cases} \qquad \text{Eq. 6}$$

This results in 3 parameters per sound stimulus: (λ_{sv}, μ_v, g_s). For the seven sounds that were not presented during the adaptation block, we took the ventriloquism effect, μ_s, equal to the trained stimulus, μ_v. We thus determined eight sound-localization gains, g_s (one for each sound type), 3

ventriloquism effects, μ_v, (one for each trained sound), and (8 x 3 =) 24 relative ventriloquism aftereffects, λ_{sv}, for each participant.

Finally, as the aftereffect gain (Eq. 3) measures a relative fraction of the ventriloquism effect, the absolute VAE, as typically reported in the literature, was determined by:

$$VAE_{sv} = \mu_v \cdot \lambda_{sv} \cdot 100\% \qquad \text{Eq. 7}$$

Statistical testing

We report on the effect sizes of the fitted parameters by determining their mean across participants. We also determined the mean difference in parameters between relevant conditions (e.g. comparing localization gain g of each narrowband sound with the gain for the Gaussian white noise). As a measure of uncertainty on both point estimates, we determined the 95% confidence intervals via bootstrapping. The means and 95% confidence intervals are reported above as: *mean* [95%*CI lower bound*; *upper bound*]. Key findings were visualized by presenting the data points, the mean and the 95% confidence interval (Ho et al., 2019).

Results

Example sound localization before, during and after exposure to visual distractor

All participants demonstrated the ventriloquism effect during the exposure block. An example of the results of the pre-, peri-, and post-adaptation phases is shown in Fig. 4 for participant S5. Figure 4A shows the stimulus-response relations of S5 for the ventriloquism sound (center frequency *v* = 5040 Hz) for the pre-, peri- and post-exposure block.

Before the exposure (Fig. 4A, blue), this participant localized the 5040-Hz sound with a considerable overshoot, yielding a fitted gain of g_v=1.4 (eqn. 6). During the exposure phase (Fig. 4A, yellow), the responses of S5 to this sound were strongly biased towards the visual distracter cue (Fig. 4A, black), yielding a ventriloquism-effect fit of μ_v=0.77. After the exposure phase (Fig.

4A, red), the responses remained biased to the - now absent - visual distracter, yielding a relative aftereffect fit of λ_{vv}=0.49.

For two other sounds in this experiment shown in Fig. 4B (500 Hz) and 4C (GWN), the localization gain was approximately equal to the gain for the exposed sound. The aftereffect for the low-frequency sound, (v=500 Hz, λ_{sv}=0.59; Fig. 4B) and for the broadband sound (v=GWN, λ_{sv}=0.53; Fig. 4C) were similar to the localization bias for the high-frequency sound exposed to an audiovisual mismatch during the exposure phase (Fig. 4A). These examples suggest an aftereffect transfer across frequencies.

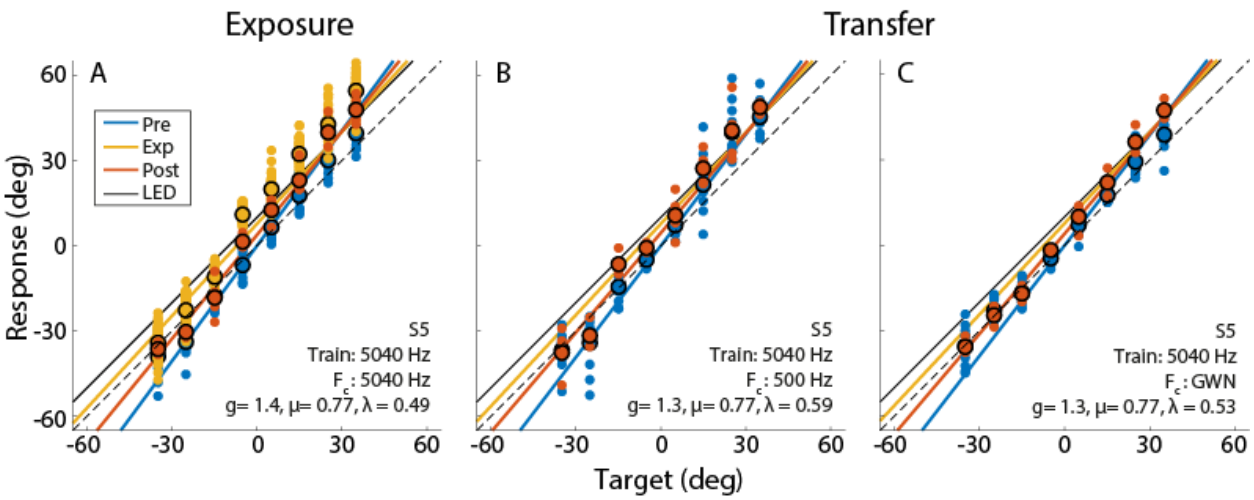

Fig. 4. Example sound localization before, during and after exposure to an audiovisual mismatch. *Stimulus-response relations of S5 in the pre-, peri-, and post-adaptation blocks.* ***(A)*** *Data for the trained exposure sound with a center frequency of 5040 Hz. The black line indicates the position of the distracter at exposure, during which responses (yellow) shifted toward the visual stimulus. The pre-adaptation data (blue) yielded a gain of 1.4 (compare Fig. 2). The post-adaptation data (red) also shifted in the direction of the ventriloquism effect.* ***(B,C)*** *The VAE is also observed for an untrained* ***(B)*** *500-Hz and* ***(C)*** *GWN sound.*

Sound localization

Figure 5 shows the pre-adaptation sound-localization gains for the eight acoustic stimuli: seven band-limited noises (center frequencies indicated) and one broadband GWN. The average gain (across all participants and sounds; Fig. 5 upper axis) was 1.25 [95.0%CI 1.17, 1.36], clearly differing from the ideal value of 1. This suggests that many of the sounds were not localized veridically (cf. Fig. 3B and Fig.4A, blue line). The gains depended slightly on the sound's frequency band. The 2000-Hz sounds tended to have the lowest gain, with a paired mean difference between the GWN and 2000-Hz sounds being -0.136 [95.0%CI -0.278, -0.0237]. Higher and lower frequency-band sounds elicited higher gains, with for example the paired mean difference between the GWN and 500-Hz sounds being 0.294 [95.0%CI 0.113, 0.464].

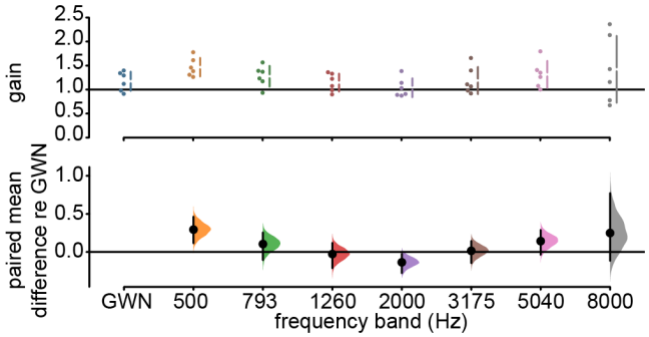

Fig. 5. Sound-localization gain. *The raw sound-localization gains, g, of each sound type in the pre-exposure block for each participant (single dot) is plotted on the upper axes. On the lower axes, the paired mean differences are plotted as bootstrap sampling distributions, for 7 comparisons against the shared control GWN are shown. Each mean difference is depicted as a black dot. Each 95% confidence interval is indicated by the ends of the black vertical error bars.*

Ventriloquism Effect

Consistent exposure of auditory targets and visual distracters evoked a strong average ventriloquism effect (Fig. 6) of 0.83 [95.0%CI 0.72, 0.94] when pooled across all three sounds. Variability of the ventriloquism effect was considerable and to a large degree idiosyncratic (Fig.

6, top) with some participants exhibiting a localization shift towards the visual distractor of near 100%, while the ventriloquism effect in others could be as low as 30%. Note that we even needed to exclude two participants with a negligible ventriloquism effect (as mentioned in Methods, section Participants). Nevertheless, the narrow-band sounds tended to evoke a stronger ventriloquism effect [μ_v=0.89 both for 793 Hz and 5040 Hz] than the GWN stimulus [μ_v=0.71]. The paired mean difference between GWN and the narrow-band sound localization gains combined was 0.18 [95.0%CI 0.06, 0.31].

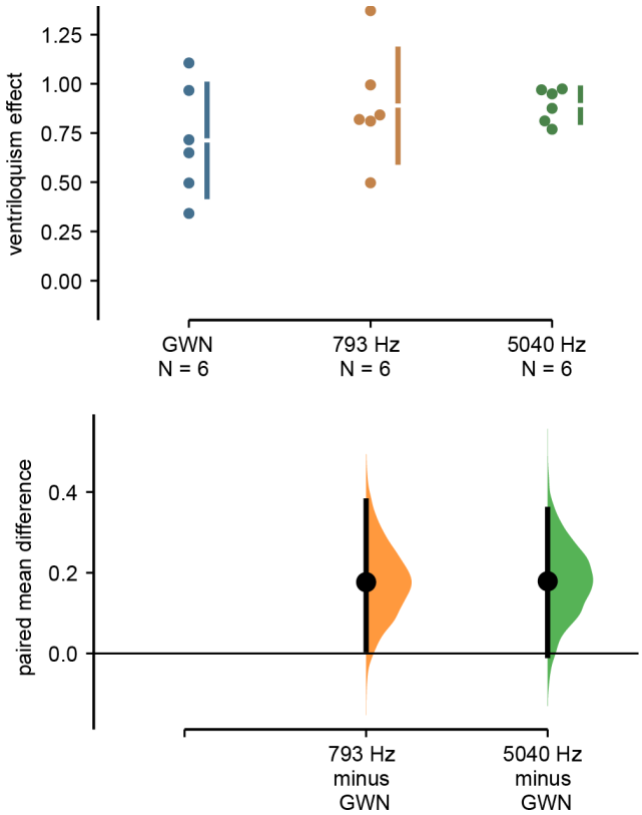

Fig. 6. Ventriloquism effect. *The ventriloquism effects, μ_v, of each exposed sound type in the peri-exposure block for each participant (single dot) is plotted on the upper axes. On the lower axes, the paired mean differences, plotted as bootstrap sampling distributions, for 2 comparisons against the shared control GWN are shown. Each mean difference is depicted as a black dot. Each 95% confidence interval is indicated by the ends of the black vertical error bars.*

Ventriloquism Aftereffect

The VAE depended on the frequency band of the eight test sounds but seemed reproducible for the three sounds (793 Hz, 5040 Hz, and GWN) presented during the exposure phase (Fig. 7A). On average, the strongest VAE of around 0.6 was obtained for the 2-kHz sound, whereas the VAE for the other frequencies was between about 0.2 and 0.4. Despite this variability, the trends seemed to suggest that the VAE was highly generalizable across the three ventriloquist sounds, at least qualitatively. To quantify this, we first corrected for this test-sound specific VAE by determining the difference between the VAEs of the narrowband and the GWN ventriloquist sounds (Fig. 7B, C). Using the corrected VAE, there was no clear difference between the trained sounds and the test sounds. The paired mean difference between the trained 793 Hz sound and any other sound (Fig. 7B) was between -0.18 (for the 500-Hz test sound) and 0.06 (For the 8000-Hz test sound). The paired mean difference between the trained 5040 Hz sound and any other sound (Fig. 7C) was between 0.00 (for the 2000-Hz test sound) and 0.19 (for the trained 793-Hz test sound). The 95% confidence intervals always included 0. This suggests that the ventriloquism aftereffect was not specific for the trained frequency (cf. Fig. 1A), which is corroborated by grouping the trained-sounds and test-sounds data: the paired mean difference between the groups is 0.019 [95.0%CI -0.0706, 0.118]. Similarly, the ventriloquism aftereffect did not seem cue-specific (cf. Fig. 1B): when grouping the test sounds according to low and high-frequency sounds, the paired mean difference between trained and other groups is 0.0865 [95.0%CI -0.0118, 0.179]. Overall, the small observed differences in VAE (Fig. 7B,C, lower axes) suggest a highly generalizable ventriloquism aftereffect in line with the aftereffect occurring at a multimodal space map (cf. Fig. 1C).

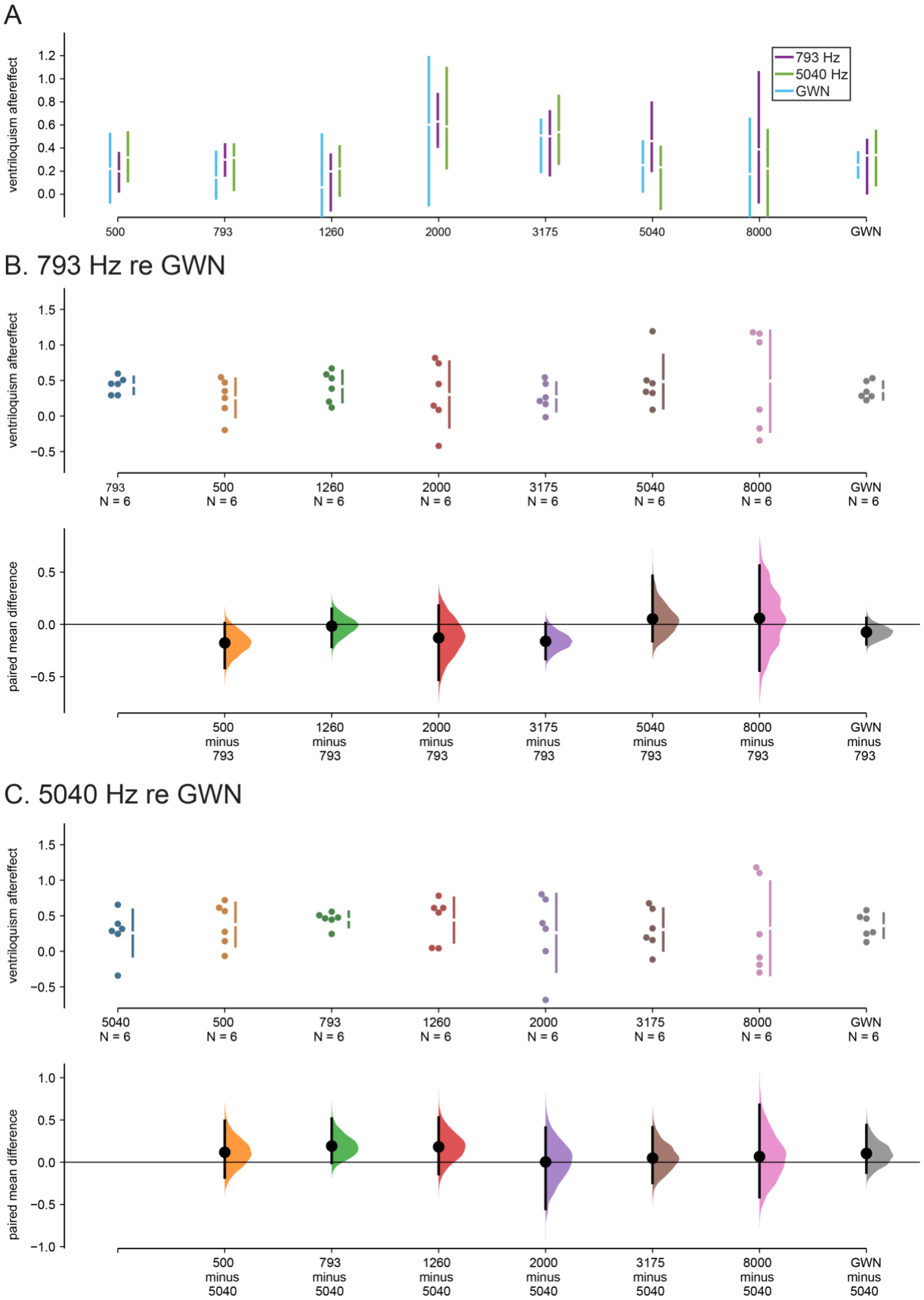
A
ventriloquism aftereffect
1.2
1.0
0.8
0.6
0.4
0.2
0.0
793 Hz
5040 Hz
GWN
500
793
1260
2000
3175
5040
8000
GWN
B. 793 Hz re GWN
ventriloquism aftereffect
1.5
1.0
0.5
0.0
−0.5
793 N = 6
500 N = 6
1260 N = 6
2000 N = 6
3175 N = 6
5040 N = 6
8000 N = 6
GWN N = 6
paired mean difference
0.5
0.0
−0.5
500 minus 793
1260 minus 793
2000 minus 793
3175 minus 793
5040 minus 793
8000 minus 793
GWN minus 793
C. 5040 Hz re GWN
ventriloquism aftereffect
1.5
1.0
0.5
0.0
−0.5
5040 N = 6
500 N = 6
793 N = 6
1260 N = 6
2000 N = 6
3175 N = 6
8000 N = 6
GWN N = 6
paired mean difference
1.0
0.5
0.0
−0.5
−1.0
500 minus 5040
793 minus 5040
1260 minus 5040
2000 minus 5040
3175 minus 5040
8000 minus 5040
GWN minus 5040

Fig. 7. Ventriloquism Aftereffect. ***A)*** *The VAE, λ, varies considerable across the 8 post-exposure frequencies (along the abscissa) but generalizes across the 3 peri-exposure sounds (indicated by vertical lines), with the strongest VAE near 2-3 kHz.* ***B)*** *The ventriloquism aftereffects for the exposure sound of 793 Hz, corrected for the GWN VAE, are plotted for each participant (single dot) on the upper axes. The correction intended to remove the frequency-dependent effects seen in A, that are independent of the sound presented during the exposure phase, while maintaining the general effect size. To that end, the VAE of GWN was removed for each peri-exposure sound and participants, and the grand average GWN VAE was added. On the lower axes, the paired mean differences, plotted as bootstrap sampling distributions, for 7 comparisons against the shared control 793 Hz are shown. Each mean difference is depicted as a black dot. Each 95% confidence interval is indicated by the ends of the black vertical error bars.* ***C)*** *Same as in B, but for the 5040-Hz sound. Note that the VAE for all sounds was similar.*

Discussion

The present study supports the hypothesis that the ventriloquism effect and aftereffect reveal multisensory neural interactions at a level in which auditory and visual stimuli are represented in a common spatial representation. Although a similar hypothesis has been forwarded before (e.g., Frissen et al., 2003, 2012; Kopco et al., 2009), our experiments have added several novel aspects, to better substantiate this claim. In what follows, we will discuss these novel features, and compare our approach and results to earlier studies.

Comparison with earlier studies

Most previous studies quantified the VE and VAE for a limited number of frequencies, not allowing one to dissociate, for example, the cue-specific hypothesis (Fig. 1B) from the spatial hypothesis (Fig. 1C). In our experiments, we increased the number of test sounds to seven

narrow-band noises with center frequencies over a wide range, from 0.5-8 kHz, and one broadband GWN. This allowed for a better quantification of the VAE across frequencies (testing generalization), as compared to earlier studies. Further, we applied the full repertoire of pre- and post-expsoure localization tests for three different VE inductors: (i) a low-frequency narrow-band sound, to specifically prime the ITD pathway in the auditory system, (ii) a high-frequency narrow-band sound that aimed for the ILD pathway, and (iii) a broadband GWN sound that stimulated all sound-localization pathways. Finally, during the audio-visual training block, participants made repeated head-movements towards the perceived sound location, which could further strengthen the potential learning effects, by invoking the full perception-action cycle (Van Opstal, 2016). Moreover, this response method allowed us to quantify and predict the stimulus-response relationships for the different stimuli and stimulus conditions across different blocks, and to analyze the responses with a quantitative regression model (Eqns. 1-6).

Non-veridical sound localization

As explained in the Introduction, the perceived disparity between the auditory target and visual distracter may deviate substantially and systematically from the physical disparity between the stimuli, and hence the putative VE and VAE will be quite different when this effect is not accounted for (Fig. 3).

Indeed, in line with earlier tone-localization studies (Musicant and Butler, 1984; Middlebrooks, 1992; Hofman and Van Opstal, 1998; Goossens and Van Opstal, 1999; Van Grootel et al., 2011), our localization data revealed considerable deviations of the perceived sound locations from the veridical, physical source locations for most narrow-band stimuli, in a frequency-dependent way (Fig. 5). By accounting for the resulting location-dependent modulation in the perceived sensory disparity (Eqns. 1-6), the three ventriloquism experiments (with v=793 Hz, 5040 Hz, and GWN) yielded remarkably similar results for the estimated VAE (Fig. 7).

Previous audiovisual-integration studies, ranging from 1970 to 2012 (Fig. 8), have reported widely different values for the absolute strength of the VAE, ranging from -13% (a ‘repelling’ effect;

Frissen et al., 2012) to nearly 100% (full transfer of the frequency-specific ventriloquism effect; Recanzone, 1998). We collected the reported values from the literature, calculated the effect size, and determined an absolute VAE median effect size of about 20% (red dashed line). The origin of this large variability is unclear. Potential factors that could have affected the reported values for the VAE are for example (i) the frequency content of the ventriloquism inductor and subsequent test sound; (ii) analysis of the data, by not accounting for non-veridical localization responses, and hence assuming a constant VE (like illustrated in Fig. 3A); (iii) use of different pointing methods to indicate the percept. The mean results from the three ventriloquism experiments in the present study are indicated in Fig. 8 at the 2021 entry (low frequency: purple, high frequency: green, GWN: blue), and lie close to the median.

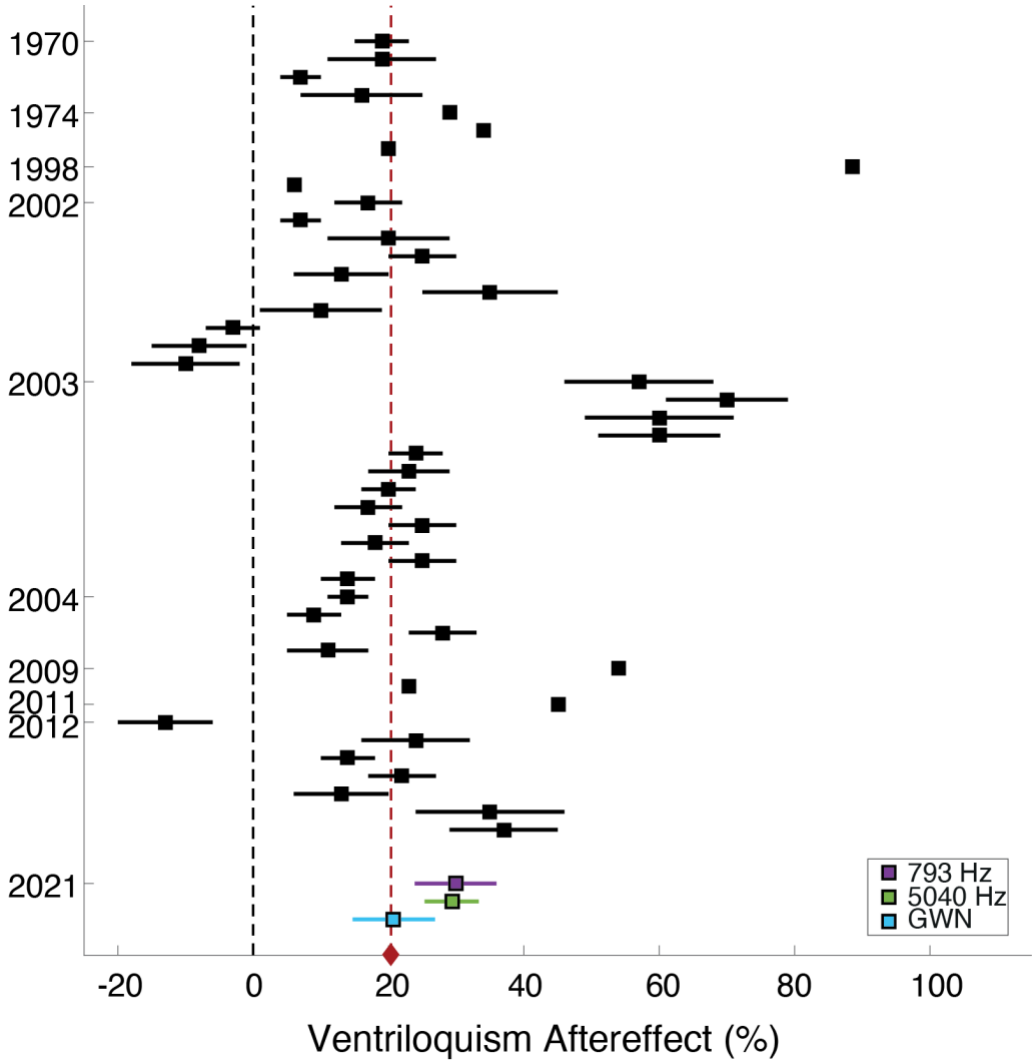

Fig. 8. Effect size of the ventriloquism aftereffect. *As reported in the literature between 1970-2012 (Recanzone, 1998; Frissen et al., 2003, 2012; Radeau and Bertleson, 1974; Canon, 1970; Kopco et al., 2009; Woods and Recanzone, 2004; Lewald, 2002; Strelnikov et al., 2011), including the current data (2021). Red diamond: median effect size derived from the previous studies. The median strength of the VAE is close to 20%. Our results for the three training sounds are included at the bottom of the figure (2021) and correspond closely to the mode.*

Behavior of the VE

All participants showed a strong ventriloquism effect for each of the three training sounds (Fig. 6). The broadband GWN induced the weakest VE when compared to the two narrow-band sounds. A possible explanation for this phenomenon could lie in the reliability of the sound-localization cues for the three different sound types. The narrow-band sounds only provide cue-information on either the ITDs (center frequency 793 Hz) or the ILDs (center frequency 5040 Hz). Although listeners can localize such sounds in the horizontal plane, the response gains deviated substantially from unity (by about 30%; Fig. 5) and, in addition, the auditory system cannot resolve the so-called 'cone of confusion' for these sounds (Blauert, 1997), as their spectral elevation cues will be highly unreliable (Musicant and Butler, 1984; Middlebrooks, 1992; Hofman and Van Opstal, 1998; Goossens and Van Opstal, 1999; Van Grootel et al., 2011; Ege et al., 2018). Thus, the two-dimensional spatial percept of these sounds will be far less certain than for the broadband GWN sound, which can be accurately localized both in azimuth and in elevation. If audio-visual integration relies on a mechanism that weighs the reliabilities of the spatial cues for the sound and visual distracter (Alais and Burr, 2004), it may be expected that the narrow-band sounds will be associated with lower integration weights with respect to the visual stimulus than the GWN. As a result, the audio-visual percept will be dominated more by the visual distracter in case of narrow-band sounds, and less by the distracter for GWN, yielding a weaker VE for the latter. Such a mechanism would reflect a maximum-likelihood estimation strategy, or more generally, a Bayesian estimation in case of a nonuniform prior. In these models, the central nervous system treats sensory inputs and prior assumptions as statistical probability distributions, where their means and variances determine the outcome of the weighting process (Ernst and Banks, 2002; Alais and Burr, 2004; Ege et al., 2018).

Behavior of the VAE

Our experiments indicate that the ventriloquism aftereffect preserves about 40-50% of the ventriloquism effect (i.e., the relative effect, λ_{sv}~0.4-0.5), generalizes across frequencies, and that

its strength varies with frequency. The VAE did not depend on the training frequency (low, high, or broadband), as the three curves in Fig. 7A are virtually indistinguishable. These results strongly support the hypothesis that the ventriloquism aftereffect reflects audio-visual interactions at a common multisensory stage where the sound-localization cues have already been processed into their spatial coordinates and represent the perceived location of the auditory stimulus (Fig. 1C).

Rather than a flat frequency response (as assumed in the simple spatial model of Fig. 1C), we found that the VAE appears to be strongest around 2-3 kHz. We conjecture that this spectral dependency could be due to a similar Bayesian cue-weighting mechanism as described above. The ITDs are most reliable for sound frequencies up to about 1.5 kHz, as for higher frequencies the interaural time delay may contain multiple periods of the sound wave, leading to phase-difference ambiguities. This ambiguity can only be resolved by relying on additional spectral input, or on other senses, such as vision. Similarly, the ILDs are most reliable for high frequencies (above 3 kHz), as the head functions effectively as a low-pass acoustic filter with a cut-off around 2 kHz. For lower frequencies, the ILDs are very weak (1 dB, or less). Thus, the frequency band between 1.5 – 3 kHz provides relatively poor localization cues for either binaural mechanism (Mills, 1958), and will likely lead to a less reliable location percept in azimuth. Following the argument described above, these sounds would therefore be more vulnerable for ventriloquism, and could thus also induce a stronger VAE.

Taken together, our data support the hypothesis, proposed by Frissen et al. (2003, 2005), that the VAE may be due to neural interactions within a common multisensory spatial representation (like a topographic map; Fig. 1C), rather than hinting at a frequency-specific interaction within the tonotopic representations of the auditory system (like in early auditory stages, or in primary auditory cortex; Recanzone, 1998). Our data add interesting aspects to these earlier reports, in that these effects may also incorporate non-acoustic information, like priors relating to the reliability of the spatial localization cues (Ege et al., 2018; Zonooz et al., 2018). Thus, the weaker the spatial cue, the stronger the ventriloquism effect and resulting aftereffect. The involvement

of such higher-order, non-acoustic information sources in the VAE further support the idea of interactions within higher multisensory integration stages.

References

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

Berger, C.C., Ehrsson, H.H. (2018) Mental Imagery Induces Cross-Modal Sensory Plasticity and Changes Future Auditory Perception. *Psychol. Sci.*, **29**, 926–935.

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

Bruns, P., Röder, B. (2015) Sensory recalibration integrates information from the immediate and the cumulative past. *Sci. Rep.*, **5**, 12739.

Bruns, P., Röder, B. (2017) Spatial and frequency specificity of the ventriloquism aftereffect revisited. *Psychol. Res.*, 1–16.

Canon, L.K. (1970) Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *J. Exp. Psychol.*, **84**, 141–147.

Corneil, B.D., Van Wanrooij, M.M., Munoz, D.P., Van Opstal, A.J. (2002) Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J. Neurophysiol.* **88**: 438-454

Ege, R., Van Opstal, A.J., Van Wanrooij, M.M. (2018) Accuracy-precision trade-off in human sound localisation. *Sci Reports* **8**: 16399

Ernst, M.O., Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429-433

Frissen, I., Vroomen, J., de Gelder, B., Bertelson, P. (2003) The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychol.* **113**, 315–327

Frissen, I., Vroomen, J., de Gelder, B., Bertelson, P. (2005) The aftereffects of ventriloquism: generalization across sound frequencies. *Acta Psychol.* **118**, 93–100

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

Tumkaya T, Aryal S, Choi H, Claridge-Chang A. Moving beyond P values: data analysis with

estimation graphics. *Nat Methods* **16**: 565–566, 2019.

Jack, C.E. & Thurlow, W.R. (1973) Effects of degree of visual association and angle of displacement on the ventriloquism effect. *Percept. Mot. Skills*, **37**, 967–979.

Kopco, N., Lin, I.-F., Shinn-Cunningham, B.G., Groh, J.M. (2009) Reference frame of the ventriloquism aftereffect. *J. Neurosci.*, **29**, 13809–13814.

Lewald, J. (2002). Rapid adaptation to auditory-visual spatial disparity. *Learning & Memory* **9:** 268-278.

Magosso, E., Cona, F., Ursino, M. (2013) A neural network model can explain ventriloquism aftereffect and its generalization across sound frequencies. *Biomed Res. Int.*, **2013**, 475427.

Magosso, E., Cuppini, C., Ursino, M. (2012) A Neural Network Model of Ventriloquism Effect and Aftereffect. *PLoS One*, **7**, e42503.

Middlebrooks, J. C. (1992) Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.* **61,** 2607–2624

Mills, A.W. (1958) On the minimum audible angle. *J Acoust Soc Am* **30**: 237-246

Moore, B.C.J. (2013) *"An Introduction to the Psychology of Hearing"*. 6th ed. Brill Publishers, Leiden, Netherlands

Musicant, A. D., Butler, R.A. (1984) The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am.* **75**, 1195–1200.

Radeau, M., Bertelson, P. (1974) Aftereffects of ventriloquism. *Quart J Exp Psychol 26: 63-71.*

Recanzone, G.H. (1998) Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proc. Natl. Acad. Sci.*, **95**, 869–875.

Robinson, D.A. (1963) A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field. *IEEE Trans BME* **10**, 137–145

Schooneveldt, G.P., Moore, B.C.J. (1989) Comulation masking release as a function of masker bandwidth, modulator bandwidth and signal duration. *J Acoust Soc Am* 85: 273-281

Strelnikov, K., Rosito, M., Barone, P. (2011). Effect of Audiovisual Training on Monaural Spatial Hearing in Horizontal Plane. *PloS One* **6**, e18344.

Van Grootel, T.J., Van Wanrooij, M.M., Van Opstal, A.J. (2011) Influence of static eye and head position on tone-evoked gaze shifts. *J. Neurosci.* 31: 17496-17504

Van Opstal, A.J. (2016) *"The auditory system and human sound-localization behavior",* Academic Press, Elsevier, Amsterdam, The Netherlands

Woods, T.M., Recanzone, G.H. (2004). Visually induced plasticity of auditory spatial perception in macaques. Current Biology **14**: 1559-1564.

Zierul, B., Röder, B., Tempelmann, C., Bruns, P., Noesselt, T. (2017) The role of auditory cortex in the spatial ventriloquism aftereffect. *Neuroimage*, **162**, 257–268.

Zonooz, B., Arani, E., Körding, K.P., Aalbers, P.A.T.R., Celikel, T., Van Opstal, A.J. *(2019)* Spectral weighting underlies perceived sound elevation. *Sci Reports* ***9****(1):1642*