A model of the reference frame of the ventriloquism aftereffect

Peter Lokša[1], Norbert Kopčo[1]

[1]Institute of Computer Science, Faculty of Science, P. J. Šafárik University, Jesenná 5, 04001 Košice, Slovakia

peter.loksa@upjs.sk, norbert.kopco@upjs.sk,

Running Title: Reference frame of Ventriloquism

ABSTRACT

Background: Ventriloquism aftereffect (VAE), observed as a shift in the perceived locations of sounds after audiovisual stimulation, requires reference frame (RF) alignment since hearing and vision encode space in different RFs (head-centered, HC, vs. eye-centered, EC). Experimental studies examining the RF of VAE found inconsistent results: a mixture of HC and EC RFs was observed for VAE induced in the central region, while a predominantly HC RF was observed in the periphery. Here, a computational model examines these inconsistencies, as well as a newly observed EC adaptation induced by AV-aligned audiovisual stimuli.

Methods: The model has two versions, each containing two additively combined components: a saccade-related component characterizing the adaptation in auditory-saccade responses, and auditory space representation adapted by ventriloquism signals either in the HC RF (HC version) or in a combination of HC and EC RFs (HEC version).

Results: The HEC model performed better than the HC model in the main simulation considering all the data, while the HC model was more appropriate when only the AV-aligned adaptation data were simulated.

Conclusion: Visual signals in a uniform mixed HC+EC RF are likely used to calibrate the auditory spatial representation, even after the EC-referenced auditory-saccade adaptation is accounted for.

1. Introduction

Auditory spatial perception is highly adaptive and visual signals often guide this adaptation. In the "ventriloquism aftereffect" (VAE), the perceived location of sounds presented alone is shifted after repeated presentations of spatially mismatched visual and auditory stimuli [1-3]. Complex transformations of spatial representations in the brain are necessary for the visual calibration of auditory space to function correctly, as visual and auditory spatial representations differ in many important ways. Here, we propose a computational model and perform a behavioral data analysis to examine the visually guided adaptation of auditory spatial representation in VAE and the related transformations of the reference frames (RFs) of auditory and visual spatial encoding.

Several previous models were developed to describe the ventriloquism aftereffect in humans and birds. The bird models examined VAE in the barn owls [4, 5] which cannot move their eyes and therefore do not need to re-align the auditory and visual RFs. The human models mainly focused on spatial and temporal aspects of the ventriloquism aftereffect [6-8], not considering the differing RFs. There are models of the audio-visual reference frame alignment, but those only consider audio-visual integration [9] and multi-sensory integration [10] when in the auditory and stimuli are presented simultaneously, like in the ventriloquism effect, not the adaptation and transformations underlying VAE.

Here, we primarily examine the reference frame (RF) in which VAE occurs. While visual space is initially encoded relative to the direction of the eye gaze, the cues for auditory space are computed relative to the orientation of the head [11]. A means of aligning these RFs is necessary by the stage at which the visual signals guide auditory spatial adaptation. Our previous studies suggest that a mixture of eye-centered and head-centered RFs is associated with recalibration in the central region of the audiovisual field [12] while the head-centered

RF dominates for VAE locally induced in a single hemifield in the visual periphery [13]. These results imply that the RF used in VAE is location dependent, possibly due to non-homogeneity in the auditory spatial representation. Specifically, recent evidence suggests that, in mammals, auditory space encoding is based on two or more spatial channels roughly aligned with the left and right hemifields of the horizontal plane [14, 15]. The current modeling explores an alternative hypothesis about the location-dependence of the RF of VAE. It assumes that the RF transformations are the same across the audio-visual field, and that the observed location-dependence is due to other adaptive processes, e.g., related to auditory saccade adaptation, as saccades were used to measure behavioral responses in the Kopco et al. [12, 13] studies. The main modeling goal is then to determine whether such a uniform, location-independent spatial adaptation is only driven by head-orientation referenced visual signals, or whether signals in eye-centered RF also contribute.

The second question explored here is how to separate the effect of auditory saccade adaptation from the ventriloquism-induced auditory space adaptation. Previous studies show that auditory saccades can overestimate or underestimate the actual sound locations [16] and that the amount of visually induced adaptation does not depend on whether the resulting saccades are hypometric or hypermetric [17]. Here, in the Appendix, we analyze the data from Kopco et al. [12, 13] to determine whether the ventriloquism effect and aftereffect show asymmetries depending on the resulting adaptation type (hypometric vs. hypermetric), as well as on the saccade amplitude magnitude. Based on this analysis, the current model assumes that the magnitude of the ventriloquism aftereffect is proportional to the magnitude of the ventriloquism effect, independent of whether these shifts result in hypometric or hypermetric saccades, and independent on the saccade magnitude.

Finally, Kopco et al. [13] observed a new adaptive phenomenon induced by aligned audiovisual stimuli presented in the periphery, exhibited as a shift in responses to sounds

presented alone in the central region. The shift magnitude depended on the gaze direction and, thus, was at least partly in the eye-centered RF. However, no such shift was observed when aligned audiovisual stimuli were presented in the central region [12]. The current model proposes a mechanism of a priori biases in the saccade responses, possibly due to auditory saccade adaptation, that can describe this phenomenon.

In the paper, we first summarize the Kopco et al. [12, 13] data modeled here, and, in the Appendix, provide a new analysis of these data to examine 1) how VAE magnitude depends on whether it results in hypometric vs. hypermetric saccades, and 2) how the VAE magnitude relates to the magnitude of the ventriloquism effect. Then, the model is introduced and two versions of it are examined in 4 simulations, each focusing on different aspects of the data and model components. The main result of the simulations is that a common location-independent mechanism can describe the data best when visual signals adapt the auditory spatial map in both head-centered and eye-centered reference frames, consistent with the idea that the reference frame of ventriloquism aftereffect is mixed.

2. Experimental data

This section summarizes the experimental methods and results from Kopco et al. [12, 13]. Additionally, Appendix presents results of a new analysis of the data aimed at examining the dependence of the results on the properties of auditory saccades used by subjects for responding.

In the experiments, ventriloquism was induced by audio-visual training trials either in the central or peripheral subregion of the horizontal audio-visual field while the eyes fixated one location (red '+' symbol; upper and middle panels of Figure 1(A)). The aftereffect was evaluated on interleaved auditory-only probe trials using a wide range of target locations while the eyes fixated one of two locations (lower panel of Figure 1(A)). The listener's task in

both audio-visual and auditory-only trials was to perform a saccade to the perceived location of the auditory stimulus/component from the FP. It was expected that the AV stimuli with displaced visual component would induce a local ventriloquism aftereffect when measured with the eyes fixating the training FP (red dash-dotted lines in Figure 1(B) illustrate this prediction for the peripheral-training experiment). Confirming this expectation, the red solid and dashed lines in Figure 1(B) show that maximum ventriloquism was induced in the peripheral and central training subregion, respectively. The critical manipulation of these experiments was that a subset of probe trials was performed with eyes fixating a new, non-training fixation point (blue '+' symbol), located 23.5° to the left of the training fixation. As illustrated by the blue dash-dotted line in Figure 1(B), if the RF of VAE is purely head-centered, then moving the eyes to a new location is expected to have no effect, resulting in the same pattern of ventriloquism for the non-training and training FPs. On the other hand, if the RF is purely eye-centered, the observed pattern of induced shifts is expected to move with the eyes when the eyes are moved to a new location, as illustrated by the cyan dash-dotted line. The experimental data showed that, in the central experiment, moving the fixation resulted in a smaller ventriloquism aftereffect with the peak moving in the direction of eye gaze (blue dashed line), while in the peripheral experiment no effect of eye gaze position was observed (blue solid line). To better visualize these results, the lower panels of Figure 1(B) shows predictions and data expressed as difference between responses from training vs. non-training FPs from the respective upper panels. The head-centered RF always predicts that the effect would be identical for the two FPs. Thus, all head-centered predictions (brown lines) are always at zero. The yellow dash-dotted line shows a hypothetical prediction for eye-centered RF, obtained by subtracting the cyan from the red dash-dotted line. Similarly, the solid and dashed yellow lines show, respectively, for the peripheral and central data, the eye-centered RF predictions obtained by subtracting from the red lines the same red lines shifted 23.5° to

the left. Finally, the black solid and dotted lines show the actual differences between the respective red and blue data from the upper panels. For the central data, the black dashed line falls approximately in the middle between the head-centered and eye-centered predictions, showing a mixed nature of the RF of VAE induced in this region. On the other hand, the black solid line is always near zero, confirming that the RF of VAE induced in the periphery is predominantly head-centered. The current model aims to describe these differences by considering a uniform representation and adaptation process that guided by signals in both eye-centered and head-centered reference frames.

The results described in Figure 1(B) are based on ventriloquism aftereffect induced by visual stimuli displaced to the left or to the right of the corresponding auditory stimuli. Figure 1(C) shows the baseline data obtained in runs with auditory and visual stimuli aligned. In the central-training experiment, the responses from the two FPs were similar, unbiased at the central locations and with a slight expansive bias in the periphery (both red and blue dotted lines are near zero in the center, negative in the left-hand portion and positive in the right-hand portion of the graph). On the other hand, in the peripheral-training experiment the responses in the central region differed between the two fixations, where the non-training FP responses fell well below the training-FP responses (compare the red and blue solid lines). Thus, the peripheral AV-aligned stimuli induced a fixation-dependent adaptation in the auditory-only responses in the central region. The black dashed and solid lines in Figure 1(C), showing the difference between the corresponding training and non-training FP data, highlight the FP-dependence of the peripheral experiment in contrast to the FP-independence in the central experiment. The current model assumes that these adaptive effects can be explained by a combination of biases in visual saccades to auditory stimuli and a visually guided adaptation in the spatial auditory representation.

---------- Insert Figure 1 around here ----------

3. Model Description

3.1 Overview

Figure 2A shows the outline of the model. The model predicts the azimuthal bias in the saccade response to an auditory-only probe (the "Response" block in panel A) as a function of the probe azimuth, with additional parameters of the fixation location on a given trial ("Probe stimulus and fixation" block) and the audio-visual training locations and the measured audio-visual response biases in a given experimental training session ("Ventriloquism" block). Thus, the model does not require information about the direction of audio-visual stimulus displacement during training (whether the visual stimuli were shifted to the left, right, or aligned with the auditory stimuli). Instead, it only uses the information about where the training occurred and what the resulting ventriloquism effect was. Here, the model assumes that there is a direct relation between the observed ventriloquism effect and aftereffect, as shown in the Appendix. The ventriloquism aftereffect prediction is then modeled as an additive combination of two components, a saccade-related bias in eye-centered reference frame and a saccade-independent visually guided adaptation of auditory space representation (square blocks in panel A). The saccade-related bias is present *a priori* and it is not directly adapted by ventriloquism, while the auditory spatial representation is locally adapted by the ventriloquism signals in different reference frames and its size also depends on the saccade-related bias.

Two versions of the model are evaluated, differing only by the assumed form of adaptation of the auditory space representation. First, in the HC model, the visual signals adapt the auditory spatial representation exclusively in the head-centered reference frame (the "HC" arrow in panel A), so the signals are assumed to be transformed to HC before inducing adaptation. In the HEC model, the visual signals adapt the auditory spatial representation in

both head-centered and eye-centered RFs ("HC" and "EC" arrows) such that the relative contribution of the HC and EC RFs can be arbitrary. I.e., the HEC model reduces to the HC model if the weight of the EC path is set to zero, or it can produce predictions using only EC RF if the HC weight is set to zero.

In summary, both models assume that the spatial representations and adaptations are uniform, predicting the same results independent of whether the training occurs in the center or in the periphery. The main difference between the two models is that the HC model assumes that the auditory space adaptation occurs purely in head-center coordinates, while it is the gaze-direction-referenced properties of the auditory saccades that cause any eye-centered effects observed in the data. On the other hand, the HEC model assumes that, even after accounting for the saccade-related effects, the auditory spatial representation receives the adaptive visual signals in both reference frames, causing adaptation that always depends on the position of the stimuli relative to the eye gaze direction. Importantly, the model assumes that if the ventriloquism aftereffect is not induced and measured by auditory saccades, as used in the Kopco et al. [12, 13]. studies, the saccade-related bias would not affect the performance.

3.2 Detailed Specification

The following model specification applies to the more general HEC model version, with the differences applying to the HC model described as needed. Panels B-D of Figure 2 provide visualizations of the behavior of different parts of the model.

Equation 1 describes the predicted bias in responses $\hat{r}$ to a given auditory stimulus location s as a weighted sum of a saccade-related bias r_E and a ventriloquism-related adaptation in auditory spatial representation r_V

$$\hat{r}(s) = r_E(s) + w \cdot r_V(s), \quad (1)$$

where $w \in [0, \infty]$ is a free parameter specifying the relative weight of the ventriloquism adaptation. In addition to the stimulus location s, the prediction (illustrated in Fig. 3D) also depends on the fixation point on a given trial f, on the training region specified by the training AV stimulus locations s_{AV}, and on the observed biases in AV stimulus responses at these locations r_{AV} (all variables in the units of degrees).

The saccade-related bias at a specific location x for eyes fixating the location f is modeled as a sigmoidal function

$$r_E(x) = \frac{2 \cdot h}{1 + e^{-k(x + cf)}} - 1, \quad (2)$$

where $h, k,$ and c are free parameters characterizing the sigmoid. The saccade-related bias (Figure 2B) is broad and referenced to the FP (i.e., it uses EC RF), exhibiting a combination of underestimations and overestimations commonly observed in studies of auditory saccades [9, 16, 18]. However, the specific shape of the functions used here was chosen to best fit the peripheral and central no-shift data shown in Fig. 1C. Specifically, the predictions roughly follow the values observed at each location in Fig. 1C when no audiovisual training is used at a given location (the central-experiment data for the right-most location triplet, the peripheral-experiment data for the central triplet, and data from both experiment for the left-most triplet). Thus, it is assumed that this saccade-related bias is present *a priori*, independent of the induced ventriloquism. Also, it is assumed that the bias only depends on the probe location re. FP location, which, for the current data means that the bias graphs for training and non-training FPs are symmetrical about the origin with respect to each other (blue and red lines in Fig. 2B).

The ventriloquism-driven auditory space adaptation causes bias defined at location x, for eyes fixating the location f, and for ventriloquism induced at training locations s_{AV} and resulting in AV response biases r_{AV}, as a weighted sum:

$$r_V(x) = \sum_{i=1}^{N} w_{v,i}(x) \cdot \left[r_{AV,i} - r_E\left(s_{AV,i}\right)\right], \quad (3)$$

where N is the number of training locations ($N = 3$ for the current study), i is an index through these locations, $s_{AV,i}$ is the i-th training location azimuth, and $r_{AV,i}$ is the AV response bias observed at the i-th training location. The differences $r_{AV,i} - r_s\left(s_{AV,i}\right)$ represent the disparity between the AV response biases (green diamonds in Figure 2B) and the saccade-related bias (red/blue lines in Figure 2B) at the training locations. The disparities are shown in Figure 2C by the red and blue full diamonds. $w_{v,i}(x)$ is the strength with which the disparity at the i-th training location adapts the spatial representation at the location x. In the HEC model, this value is a weighted sum of the adaptation strengths in head-centered and eye-centered reference frames, defined as:

$$w_{v,i}(x) = (1 - w_E) \cdot w_{vH,i}(x) + w_E \cdot w_{vE,i}(x), \quad (4)$$

where $w_E \in (0, 1)$ is a parameter determining the relative weight of the EC reference frame vs. the HC RF (in the HC model, $w_E = 0$). Finally, $w_{vH,i}$ and $w_{vE,i}$ use normalized Gaussian functions centered at training locations as a measure of influence of the i-th training location on the target location x, in the two reference frames:

$$w_{vH,i}(x) = G\left(x, s_{AV,i}, \sigma_H\right), \quad (5)$$

$$w_{vE,i}(x) = G\left(x, s_{AV,i} + f - 11.25°, \sigma_E\right), \quad (6)$$

$$G(x, \mu, \sigma) = \frac{\mathcal{N}(x,\mu,\sigma)}{\sum_{i=1}^{N} \mathcal{N}(7.5\cdot(i-2),0,\sigma)}, \quad (7)$$

In Eqs. 5 and 6, the parameters σ_H and σ_E represent the width of the influence of the ventriloquism shift at individual training locations, separately for the two reference frames. $w_{vH,i}$ (Eq. 5) is always centered on the i-th training location in the HC RF, whereas $w_{vE,i}$ (Eq. 6) is centered on the i-th training location in the EC RF (for the training FP, the two RFs are aligned). Finally, the Gaussian functions are normalized (Eq. 7) such that the maximum $w_{vH,i}$

or $w_{vE,i}$ after summing across the three training locations is 1 (the normalization locations $7.5 \cdot (i-2)$ are specific for the current training and they need to be modified for other data with different training locations).

Figure 2C shows the operation of the ventriloquism adaptation. As mentioned above, the red and blue filled diamonds are the disparities at the individual training locations driving the adaptation in HC RF. The blue open diamonds are identical to the blue filled diamonds except that they are shifted to the left by the difference between the two FPs to illustrate how the eye gaze shift affects where the adaptation is expected to occur in the EC RF. The red and blue lines are then the resulting biases r_V for the two fixation locations, each corresponding to the sum of Gaussians centered at different training locations in the two RFs (and with widths defined by the σ's). Parameter w_E determines the relative weights of the peaks in the blue line corresponding to the open diamonds vs. those corresponding to the filled diamonds. In summary, the blue and red lines show how visually guided adaptation is local and RF-dependent, decreasing with distance from location at which AV stimuli were present in HC and EC RFs. It also shows that since adaptation causes shifts from the saccade-bias response locations towards AV response locations, if AV responses fall on saccade bias locations, no visually guided adaptation is predicted to occur.

Finally, Figure 2D shows that the model prediction is a sum of the saccade bias (from Figure 2B) and ventriloquism bias (Figure 2C) weighted by the parameter w (note that no scaling parameter is needed for the saccade bias as parameter h already can make this bias arbitrarily large).

---------- Insert Figure 2 around here ----------

4. Methods

4.1 Stimuli

The data from studies of Kopco et al. [12, 13], simulated here, induced ventriloquism by presenting training stimuli with visual component either shifted to the left, to the right, or aligned with the auditory component, while the eyes fixated one location (Fig. 1A; upper and middle panels). The aftereffect was always measured by presenting auditory-only stimuli while eyes fixated one of the two FPs (Figure 1A; lower panel). Thus, nominally, there were 6 conditions (3 shift directions by 2 training regions), corresponding to AV locations and responses shown by triplets of open symbols in Figure A1A. For these conditions, predictions could be compared to data for 9 locations at 2 FPs. However, the main experimental results simulated here were observed when differences between FPs were considered on aftereffect magnitude data, obtained by subtracting positive-shift data from negative-shift data and halving the result (Figure 1B; lower right panel; note that the latter difference is equivalent to averaging the magnitudes of “positive shift – no shift” and “negative shift – no shift”). These “double differential” (“positive – negative” difference of “training FP – non-training FP” difference) data were the most stable as they eliminated a lot of between-subject variability related to individual biases in responses (as will be illustrated later). Therefore, to focus the model on these important differences, the data were also transformed into the difference representation in two steps.

First, the data for the two training FPs were orthogonally transformed such that instead of using training and non-training FP, a sum and a difference across the two FPs was used. I.e., instead of having for each condition 18 data points corresponding to 9 locations at 2 FPs, we used 18 data points consisting of 9 locations summed across the two FPs and 9 locations for difference across the 2 FPs.

Second, the positive-shift and negative-shift condition data were transformed in a similar way, such that instead of positive and negative shift we used the aftereffect magnitude (i.e., a halved difference between the two shifts) and average across the two shifts. The no-shift data were left unmodified.

The complete data set therefore consisted of 108 data points [9 (locations) x 2 (transformed FPs) x 3 (transformed shifts) x 2 (training regions)]. Across-subject mean and standard deviation data were used in the simulations.

4.2 Simulations

Four simulations were performed in this study, each assessing both the HC and HEC models on a different subset of the Kopco et al. [12, 13] data. The first two simulations, No-Shift and All Data simulations, tested two main hypotheses about the current data and reference frame. Two supplementary simulations, Central Data and Peripheral Data simulations, were performed confirm that the model behavior matches the conclusions of the Kopco et al. [12, 13] studies when considered separately.

No Shift simulation assessed the models on the AV-aligned baseline no-shift data from both experiments (Figure 1C), examining the interaction between the saccade-related bias and visual signals when no ventriloquism is induced.

All Data simulation is the main simulation of this study. In this simulation the models were fitted on the complete dataset from both experiments (Figure 1B and C) to examine whether a uniform representation of the reference frame of ventriloquism aftereffect is mixed or purely head-centered.

Central Data simulation fitted only the central-training data from the positive-shift and negative-shift conditions (dashed lines in Figure 1B) while predictions were generated for

all the data. The main goal was to examine the reference frame in which the ventriloquism aftereffect is induced in the central region.

Peripheral Data simulation fitted only the peripheral-training data from the positive-shift and negative-shift conditions (solid lines in Figure 1B) while predictions were generated for all the data. The main goal was to examine the reference frame in which the ventriloquism aftereffect is induced in the audiovisual periphery.

4.3 Model Fitting and Evaluation

Each simulation was performed by fitting the two models to the corresponding subset of the transformed data using a two-step procedure. First, a systematic search through the parameter space was performed, using all combinations of 10 values for each parameter, listed in Table 1 (HEC model used all 7 parameters, while HC model only used 5 of them). The limits of the range were chosen by piloting to cover the expected range of behaviors of the model. Note that quadratic spacing was chosen for parameters k and c as the behavior of the sigmoidal function varies non-uniformly with the parameter values (k was sampled more densely at the lower end of the range, c at the higher end). Then we selected the best 100 parameter combinations in terms of weighted MSE, in which each data point was weighted by the inverse of the across-subject standard deviation in that data point. These parameter combinations were then used as starting positions for non-linear iterative least-squares fitting procedure (Matlab function lsqnonlin) which, again, minimized the weighted MSE. The parameter values obtained by the best of these fits were chosen as the optimal values.

---------- Insert Table 1 around here ----------

To compare the models' performance while accounting for the number of parameters used by each model, we computed the Akaike information criterion AICc [19, 20] for each optimal fit, defined as:

$$\mathrm{AICc} = -2\log(L) + 2K + 2K\frac{K+1}{n-K-1}, \quad (8)$$

$$\log(L) = -\frac{n}{2}\left(\log(2\pi)+\log\frac{\mathrm{SSE}(X)}{n}+1\right) \quad (9)$$

where n is the number of experimental data points, K is the number of fitted parameters, and $\mathrm{SSE}(X)$ is the sum of squares of errors across the data points (i.e., differences between predictions and across-subject mean data x_{i}) weighted for each data point by the inverse of its across-subject standard deviation $\frac{1}{SD(x_j)}$. In general, the model with the lower AICc is considered to be a better fit for the data. Then, to determine whether the data provide substantial support for one model over the other one, we computed ΔAIC as the difference in AICc values of the model with the higher AICc vs. the one with the lower AICc. And, we use the following rule to determine whether the model with the lower AICc is substantially better than the other model [19]: "Models having $\Delta\mathrm{AIC} \leq 2$ have substantial support (evidence), those in which $4 \leq \Delta\mathrm{AIC} \leq 7$ have considerably less support, and models having $\Delta\mathrm{AIC} > 10$ have essentially no support.". Thus, only if ΔAIC is substantially larger than 2, the result is interpreted as evidence in favor of the model with lower AICc.

5. Simulation Hypotheses and Results

The results of the 4 simulations performed in this study are summarized in Table 2, which shows for each simulation and model the fitted model parameter values and the model's performance measured using the AICc criterium.

---------- Insert Table 2 around here ----------

364 *5.1 No-shift simulation*

365 This simulation focused on the AV aligned data, examining the hypothesis that *the*
366 *saccade-related bias combined with auditory space adaptation in HC RF causes the training-*
367 *region-dependent differences in the AV-aligned baseline data* (Figure 1C). I.e., it was
368 predicted that EC visual signals adapting the auditory space representation do not need to be
369 considered to explain the different adaptation effects observed in central vs. peripheral AV-
370 aligned training. This hypothesis would be confirmed if the two models, HC and HEC,
371 captured the behavioral data equally well.

372 Figure 3 presents the results of the simulation of the AV-aligned baseline no-shift
373 condition from both experiments. Panel A shows the biases of the two model components
374 (rows) for each of the two models (colors) with the fitted parameters as listed in Table 2,
375 separately for the two fixation points (columns). The same fitted model parameters apply to
376 both the central and peripheral training experiments. For the saccade-related bias (upper row)
377 that means that the plotted graphs apply to both data equally. However, for the auditory space
378 adaptation component (lower row), the plotted graphs apply to central training, since they
379 show the effect of training at the 3 central locations (-7.5°, 0°, +7.5°). The graphs need to be
380 shifted to the right by 22.5° to see their effect for peripheral training data.

381 Panel B shows the data (circles with error bars corresponding to the standard error of
382 the mean) and predictions of the two models (lines), separately for the two training points
383 (upper and middle rows), as well as for the difference between the FPs (lower row). The
384 columns represent the two training regions. Each prediction in the upper and middle rows is,
385 roughly, a weighted sum of the corresponding components from panel A, while the
386 predictions in the lower row of panel B show the differences of the predictions from the upper
387 and middle rows.

Considering the model predictions of the mean data, both models captured all the significant trends in these data. Specifically, for the central training data, both models predicted the slight expansion of the space for the central training data identical for both FPs (upper and middle row of the left-hand column), as well as the FP-dependence of the peripheral training data at the central locations (upper and middle row of the right-hand column). Most importantly, both models captured very well the difference data, which are near zero for the central training experiment and have a positive deviation for the peripheral training (bottom row). This conclusion is confirmed by the AICc evaluation which showed no evidence that either of the models should be preferred ($\Delta AIC = 2.4$).

The data in panel B are replotted from Figure 1C, now also including the error bars. These error bars show that there was a lot of across-subject variability when the individual FPs were considered (upper and middle row), while a large portion of that variability was eliminated when the differences in biases across the FPs were computed (lower row). This illustrates why the models were fitted on the transformed data, as those were much more consistent across subjects, and, with the transformation, the fitting weighed the difference data (lower row) more as they were much more reliable. Note that the second transformed data set, the average across FPs, is not shown, as it can be easily estimated from the individual FP data in the upper two rows of panel B.

Panel A illustrates how the models achieved the correct prediction. Both models predicted similar saccade-related bias, consisting of expansion at the peripheral target locations (+/-15°, +/-22.5°, and +/-30°) and bias towards the fixation location for the central 3 locations (upper row). This saccade-related bias was then modulated by the auditory space adaptation such that at the training locations the model predictions were shifted towards the AV responses, which were near zero for both the central and peripheral training (FigureA1A). The HC model predicts that this “corrective” ventriloquism shift only occurred in HC RF

(brown lines in the lower row of panels), while the HEC model predicts a considerable contribution of the EC RF (magenta lines at locations -30° to -15° at the bottom right). However that contribution only had a small effect on the overall predictions, as shown by the small differences between the brown and magenta lines in panel B.

---------- Insert Figure 3 around here ----------

5.2 All Data simulation

This was the main simulation of this study. The two models were fitted on the positive-shift and negative-shift data, in addition to the no-shift data from the previous simulation (Figure 1B and C). Also, the simulation was performed on the data from both experiments. Thus it assumed that the reference frame of ventriloquism aftereffect is uniform across the audiovisual field, as the models were optimized to fit both the central and peripheral training data simultaneously. The simulation further assumed that the saccade-related component of the model accounts for all the saccade-related effects (which are EC-referenced), an assumption supported by the results of the No Shift simulation. With these assumptions, the simulation examined the hypothesis that the *RF is mixed, using visual signals in both head-centered and eye-centered coordinates*. This hypothesis would be confirmed if the HEC model, using both HC and EC referenced visual signals, captured the behavioral data significantly better than the HC model, which only uses HC RF for the ventriloquism adaptation of the auditory space.

Figure 4 presents the results of this simulation. Panel A shows the biases of the two model components for the fitted parameter values from Table 2, in a format similar to panel A of Figure 3. Panel B shows the data (circles with error bars corresponding to the standard error of the mean) and predictions of the two models (lines). Panel B shows for this simulation only the difference of Training vs. Non-training FP data, equivalent to the black

lines in Figure 1B and 1C. The upper row of panel B shows the no-shift data replotted from Figure 1C (also shown in the bottom row of Figure 3B), while in the lower row shows the difference between the positive-shift and negative-shift data, equivalent to a doubling of the aftereffect magnitude data from Figure 1B (black solid and dashed lines).

The data and model predictions addressing the main hypothesis of this simulation are in the lower row of panel B. The central training data show a large positive deviation in the middle of the target range, corresponding to the mixed reference frame, while the peripheral training data are always close to zero, an evidence of the head-centered RF. The HEC model (magenta line) approximates this pattern by predicting a positive deviation in both training regions accompanied by a negative deviation of similar size for the targets to the left of the training regions. This pattern captures the main characteristics of the data even though the predicted positive deviation is weaker than that observed for the central central-training data. On the other hand, the HC model (brown line) always predicts no deviation from zero, as that model assumes that the adaptation always occurs in the HC RF. These differences between the models confirm the hypothesis that auditory representation is adapted uniformly by visual signals in both head-center and eye-center reference frames. This conclusion is confirmed by the AICc evaluation which showed almost no support for the HC model compared to the HEC model ($\Delta\mathrm{AIC} = 7.9$).

The model predictions for the no-shift data (upper row of panel B) are almost identical for the two models. Thus, the difference in performance between the models cannot be explained by differences in accounting for the no-shift data. Notably, the predictions for the two training regions are fairly similar to each other, and slightly worse than those obtained in the No Shift simulation. However, they still capture the pattern of biases fairly well. Finally, note that the predictions for the average of positive and negative shift data is not shown, even though these transformed data were also used for fitting. These data were omitted as both the

462 data and model predictions are very similar to the no-shift results shown in the upper row of
463 panel B.

464 Looking at across-subject variability in the data, the error bars in panel B tend to be
465 smaller for the positive-vs-negative shift plots (lower row) than for the no-shift plots (upper
466 row). This difference is in fact much larger, since the plotted error bars are for the difference
467 between the two shift directions, whereas the aftereffect magnitude equal to half of the
468 difference was used in the fitting. This shows that additional between-subject variability was
469 caused by idiosyncratic biases in each subject's responses that are consistent within each
470 subject, and which therefore cancel out when the difference between positive and negative
471 shift data is computed. This again shows the importance of fitting the models on the
472 transformed data, which resulted in weighing the positive-vs-negative shift difference data
473 (lower row) even more than the no-shift training-vs-non-training FP data (upper row).

474 Panel A illustrates the behavior of individual components that resulted in the models'
475 predictions. The saccade-related bias is almost identical for the two models (upper row), and
476 overall similar to the pattern observed in the NoShift simulation (Figure 3A). The auditory
477 space adaptation is broad for both models, and only slightly different between the models
478 (magenta vs. brown lines between in the lower row of Figure 4B). The size of the difference is
479 mainly determined by parameter w_E (see Table 2) which defines the relative contribution of
480 the eye-centered vs. head-centered RF to the combined representation in the HEC model (in
481 this simulation $w_E = 0.15$, indicating that the EC RF only had a 15% weight in the mixed
482 reference frame). So, it can be concluded that even though this contribution is highly
483 significant, the HC RF has still a dominant role when uniform representation of the auditory
484 space is assumed.

485 ---------- Insert Figure 4 around here ----------

5.3 Central and Peripheral Data simulations

Two additional simulations were performed, each of them fitting separately the data for only one training region. The main goal of the simulations was to verify that, when the models are fitted to the two data sets separately, they will confirm the conclusions of the behavioral experiments about the mixed reference frame for the central-training data and the head-centered reference frame for the peripheral-training data. Additionally, these simulations only fitted the transformed positive-shift and negative-shift data, while also producing model predictions for the no-shift data. Thereby, the simulations tested whether the behavior of the saccade-related model component observed in the previous simulations is dependent on the presence of the no-shift data, or whether the models would find a similar predicted pattern even if only the positive/negative shift data are considered.

Central Data simulation fitted only the central-training data from the positive-shift and negative-shift conditions (dashed lines in Figure 1B). The main hypothesis tested in the simulation was that the *RF is mixed when VAE is induced in the central region*. This hypothesis would be confirmed if the HEC model is significantly better than the HC model. Figure 5 presents the results of this simulation using a layout identical to Figure 4. The lower row of panel B shows the predictions of the two models for the difference data. As expected, the HEC model (magenta) fits the central-training data well (better than in the All Data simulation) while the HC model's prediction (brown) is again fixed at zero. This difference confirms the hypothesis that the EC RF contributes significantly to the ventriloquism adaptation in central region, a conclusion also confirmed by the AICc evaluation (HEC model better than HC model; $\Delta\mathrm{AIC} = 5.9$). However, it is also noticeable that the HEC model underestimates the central data for targets at azimuths around 0° while it predicts a negative deviation at azimuths around -20°, not observed in the data. This negative deviation is due to the structure of the model which always predicts that a positive deviation is accompanied by a

511 negative deviation at locations shifted in the direction of the new, non-training FP location.
512 For the peripheral experiment, the HEC model predictions depart considerably from the data,
513 as expected since the data do not show a strong EC RF contribution. On the other hand, for
514 the no-shift data, both models largely capture the main trends even though they were not fitted
515 on these data (upper row of panel B), confirming that the FP-dependent adaptation observed
516 in the no-shift data is not specific to these data as the model generalizes to predict it even if
517 only trained on the positive and negative shift data.

518 Considering the individual model components (Panel A), the results are overall similar
519 to the All Data simulation (Figure 4). The main difference in the current simulation is that the
520 EC-referenced contribution to auditory spatial adaptation in the HEC model is considerably
521 stronger, resulting in larger differences between the two models (bottom row). However, even
522 here the HC RF still has more weight ($w_E = 0.3$ in Table 2), suggesting that it is the more
523 dominant RF for ventriloquism aftereffect in general.

524 ---------- Insert Figure 5 around here ----------

525 Peripheral Data simulation fitted only the peripheral-training data from the positive-
526 shift and negative-shift conditions (dashed lines in Figure 1B). The main goal was to confirm
527 the hypothesis that the *RF is head-centered when VEA is induced in the peripheral region*, in
528 agreement with the behavioral results. This hypothesis would be confirmed if the HEC and
529 HC models performed similarly in the simulation.

530 Figure 6 presents the results of this simulation using a layout identical to Figure 4. The
531 lower row of panel B shows the predictions of the two models for the positive vs. negative
532 shift difference data. As expected, both models fit the near-zero peripheral-training data well,
533 while failing to predict the central-training data. This confirms that the EC RF does not
534 contribute to the ventriloquism adaptation in the peripheral region, a conclusion also
535 supported by the AICc evaluation, in which the HC model is better than the HEC model;

536 $\Delta AIC = 5.6$ in Table 2). Similar to the Central Data simulation, for the no-shift data, both
537 models largely captured the main trends even though they were not fitted on these data (upper
538 row of panel B). These results are also confirmed when considering the individual model
539 components (Panel A). First, the saccade-related bias component (upper row) again behaves
540 identically in the two models similarly to the previous simulations. Second, the auditory space
541 adaptation component (lower row) behaves nearly identically for the two models, determined
542 by the low the relative weight of the EC RF in the HEC model ($w_E = 0.04$ in Table 2).

543 ---------- Insert Figure 6 around here ----------

544 *5.4 Model parameter values*

545 The behavior of the models in different conditions can be analyzed by looking at the
546 fitted values of the model parameters. Here, the first main modeling question concerned the
547 ability of the models to predict the EC-dependence of the no-shift data observed in the
548 peripheral, but not in the central, training condition. The critical model parameters here are
549 the parameters *h* and *w*, which determine the relative strength of the saccade-related and
550 auditory space adaptation components of the model (Figure A1 and Table 2). The values of
551 the two parameters are overall similar in all simulations, suggesting that both components
552 contributed critically to all the predictions.

553 The parameter w_E determined the relative strength of the EC RF contribution to the
554 ventriloquism-driven auditory spatial adaptation, while the parameters σ_H and σ_E determined,
555 respectively, how broad-vs-specific was the influence of the HC and EC RFs. The value of
556 w_E was always much smaller than 0.5 (in relevant simulations smaller or equal to 0.3) and σ_H
557 was always much larger than σ_E. Both these observations indicate that while the EC-
558 referenced signals influence the ventriloquism adaptation significantly, their effect is mostly
559 modulatory, while the HC-referenced signals dominate.

Finally, the fitted values of parameters *k* and *c* did not change dramatically across the simulations, always resulting in similar predictions about the saccade-related bias component of the model.

6. Summary and Discussion

The HC/HEC model introduced here aims to characterize the reference frame in which auditory and visual signals are combined to induce the ventriloquism aftereffect. It focuses on the experimental data in which ventriloquism was induced locally in either the audiovisual center or periphery, in which a change in fixation point was used to dissociate the head-center from eye-centered reference frames, and in which saccades were used for responding during training and testing [12, 13]. The model assumes a population of adaptive units representing the auditory space with auditory and visual inputs, similar to the channel processing model proposed in [21]. However, instead of explicitly implementing a population of units, it describes the adaptive effects by only considering the locations from which the auditory components of audiovisual training stimuli were presented. Then, for each unit there is a Gaussian neighborhood in which the AV training affects the A-only responses in either HC-only RF (HC model) or in a combined HC+EC RF (HEC model). Also, the model assumes that there are intrinsic biases associated with auditory saccade responses, and that the effect of ventriloquism is to shift the auditory-only responses from these saccade-related biases towards the locations of the responses on the audiovisual training trials.

Since the model only uses the responses on audiovisual training trials to guide adaptation, independent of the direction of audiovisual disparity used during training, and independent of whether the adaptation results in hypometric or hypermetric saccades, it is assumed that there is a direct relation between the audiovisual responses during training and the auditory-only responses during testing. Specifically, the assumed relationship is that the

ratio of observed ventriloquism aftereffect to the observed ventriloquism effect is constant, as confirmed by our behavioral data analysis (see Appendix) which found a ratio of approximately 0.5. This ratio is not aftereffect by whether the aftereffect results in hypometric or hypermetric saccades, consistent with Pages and Groh [17]. However, the analysis also found that there is an asymmetry in the ventriloquism effect when measured using audiovisual saccades. Specifically, the effect reaches 100% of audio-visual disparity if resulting in hypometric saccades, while it is only 80% of the disparity when resulting in hypermetric saccades. Future studies will need to determine whether there is really a difference in the presence/absence of the hypo/hypermetric asymmetry when saccades are used for ventriloquism effect and aftereffect measurement, or whether the current results are different for the effect vs. aftereffect only because the aftereffect data are noisier.

The four simulations presented here showed that the HC/HEC model can describe the different phenomena observed in the Kopco et al. [12, 13] studies. First, in the No-Shift simulation, the simpler HC model accurately predicted the newly reported adaptation by AV-aligned stimuli [13] as a combination of the intrinsically present saccade-related biases locally "corrected" by the visually guided adaptation at the training locations. Thus, the model predicts that this AV-aligned adaptation for the peripheral-training data is purely driven by some adaptive processes affecting the motor representations related to audiovisual/auditory saccades. This, as well as the existence of the saccade-related bias component of the model, can be tested in future studies, as the currently available data are not consistent as to whether auditory saccades are predominantly hypermetric or hypometric [16, 18]. Both these predictions can be experimentally tested by performing ventriloquism experiments in which saccades are not used for responding [22].

The second, All Data simulation addressed the main question of this study about the reference frame of the ventriloquism aftereffect. Its results provide an evidence that a uniform

auditory spatial representation uses a mixed reference frame, with visual signals adapting the auditory spatial representation in both head-centered and eye-centered RFs, as implemented in the HEC model and consistent with physiological studies [23, 24]. Importantly, the current results suggest that, in the mixed frame, the relative contribution of the EC RF is only 15% vs. 85% for the HC RF. Moreover, even when only the central-training data are considered (Central-Data simulation), the relative contribution of the EC only reaches 30%. Thus, the HC RF is always dominant for the ventriloquism aftereffect adaptation, an observation that is further supporter by the comparison of the fitted sigma parameters (which showed that the HC-referenced adaptation is more broad than the EC-referenced adaptation). The second simulation also showed that the model in its current form always predicts the same difference in biases between the FPs, independent of the training region. This effect is mainly due to the implicit model assumption that the distribution of the spatial channels is uniform across space. If the model assumed a denser representation of space near the midline (e.g., see [25]), it could predict adaptation that is stronger in the center than in the periphery.

Importantly, the current model was fitted on data transformed so that the differences between the two FPs and differences between the positive and negative shift data were used. This was particularly critical for this simulation in which the EC contribution is visible when the double difference is computed, and it was also important since, in this representation, a lot of noise in the data is removed. Note that when the All data simulation was repeated on untransformed data, the AICc evaluation did not find a significant difference between the HC and HEC models, since the across-subject variability in the responses considered separately for the two FPs was too large, dominating over the differences between the FPs critical to evaluate the reference frames (data not shown).

The final two simulations examined the model behavior when fitted separately to the central vs. peripheral training data. In both simulations the model predictions were in

agreement with the behavioral data. Specifically, the HEC model using a mixed reference frame better predicted the central data, while the HC model using the head-centered reference frame better predicted the peripheral data. The central-data simulation also showed one weakness of the model: in its current form it always predicts that if there is a region in which VAE magnitude is larger for the training-FP than non-training-FP data, then there also has to be a region in which the relationship is reversed. An extension of the model which would make the strength of the adaptation depend not only on the distance from the training stimuli, but also on the distance from the training FP, could correct this discrepancy.

Finally, the Central and Peripheral Data simulations accurately captured the no-shift data, even though the models were not fitted on them, confirming that the pattern of adaptation exhibited in these data is also present in the positive-shift and negative-shift data from which it can generalize to the no-shift data. However, as discussed above, the no-shift data biases are most likely related to the saccade responses, not to the spatial representation adapted by ventriloquism, which is of primary interest here.

The neural mechanisms of the ventriloquism aftereffect and its reference frame are not well understood. Cortical areas involved in ventriloquism aftereffect likely include Heschl's gyrus, planum temporale, intraparietal sulcus, and inferior parietal lobule [26-29]. Multiple studies found some form of hybrid representation or mixed auditory and visual signals in several areas of the auditory pathway, including the inferior colliculus [30], primary auditory cortex [31], the posterior parietal cortex [23, 32, 33], as well as in the areas responsible for planning saccades in the superior colliculus and the frontal eye fields [34, 35]. In the current model, the saccade-related component likely corresponds to the saccade-planning areas. The auditory space representation component likely corresponds to the higher auditory cortical areas or the posterior parietal areas, not the primary cortical areas. This can be expected because there is growing evidence that, in mammals, auditory space is primarily encoded non-

659 homogeneously, based on two spatial channels roughly aligned with the left and right
660 hemifields of the horizontal plane [14, 15, 36-38] and the ventriloquism adaptations modeled
661 here are local (within a hemifield or just in the central region), not consistent with broad
662 adaptation predicted by the hemifield code. However, note that there are also theories which
663 incorporate additional channels, such as a central channel, in addition to the hemifield
664 channels [39]. Such extended models might be compatible with the current data.

665 Even though most previous recalibration studies examined the aftereffect on the time
666 scales of minutes [1, 2, 40, 41], recent studies demonstrated that it be elicited very rapidly,
667 e.g., by a single trial with audio-visual conflict [42]. If it is the case that the adaptive
668 processes underlying the ventriloquism aftereffect occur on multiple time scales, as also
669 suggested in several models of slower ventriloquism aftereffect [6, 7], then an open question
670 is whether the reference frame is the same at the different scales or whether it is different. The
671 current results are mostly applicable to the slow adaptation on the time scale of minutes, while
672 the RF on the shorter time scales has not been previously explored.

673 In summary, while some previous models considered the reference frame of the
674 ventriloquism effect [9, 10], the current HC/HEC model is, to our knowledge, the first one to
675 focus on the RF of the ventriloquism aftereffect. In addition, it also considers how saccade-
676 related adaptation might influence auditory saccades. In the future, it can be combined with
677 the existing models of spatial and temporal characteristics of the ventriloquism aftereffect to
678 obtain a more general model of this important multisensory phenomenon.

679

680 **Acknowledgments**

681 This work was supported by VEGA-1/0355/20 and VVGS UPJS VVGS-2020-1514.

682

683

References and links

685 1. Recanzone, G.H., *Rapidly induced auditory plasticity: The ventriloquism aftereffect.*
686 Proceedings of the National Academy of Sciences of the United States of America,
687 1998. 95(3): p. 869-875.

688 2. Woods, T.M. and G.H. Recanzone, *Visually Induced Plasticity of Auditory Spatial*
689 *Perception in Macaques.* Current Biology, 2004. 14: p. 1559-1564.

690 3. Bertelson, P., et al., *The aftereffects of ventriloquism: Patterns of spatial*
691 *generalization*. Perception and Psychophysics, 2006. 68(3): p. 428-436.

692 4. Haessly, A., J. Sirosh, and R. Miikkulainen. *A model of visually guided plasticity of*
693 *the auditory spatial map in the barn owl*. in *Seventeenth Annual Meetings of the*
694 *Cognitive Science Society*. 1995. Pittsburgh, PA: Erlbaum.

695 5. Oess, T., M.O. Ernst, and H. Neumann, *Computational principles of neural*
696 *adaptation for binaural signal integration.* PLOS Comput Biol, 2020. 16(7).

697 6. Bosen, A.K., et al., *Multiple time scales of the ventriloquism aftereffect.* PLoS ONE,
698 2018. 13(8).

699 7. Watson, D.M., et al., *Distinct mechanisms govern recalibration to audio-visual*
700 *discrepancies in remote and recent history.* Sci Rep, 2019. 9.

701 8. Shinn-Cunningham, B.G., N. Kopco, and T.J. Martin, *Localizing nearby sound*
702 *sources in a classroom: Binaural room impulse resonses.* Journal of the Acoustical
703 Society of America, 2005. 117(5): p. 3100-3115.

9. Razavi, B., W.E. O'Neill, and G.D. Paige, *Auditory Spatial Perception Dynamically Realigns with Changing Eye Position.* Journal of Neuroscience, 2007. 27: p. 10249-10258.

10. Pouget, A., S. Deneve, and J.R. Duhamel, *A computational perspective on the neural basis of multisensory spatial representations.* Nature Reviews Neuroscience, 2002. 3: p. 741-747.

11. Groh, J.M. and D.L. Sparks, *Two models for transforming auditory signals from head-centered to eye-centered coordinates.* Biological Cybernetics, 1992. 67: p. 291-302.

12. Kopco, N., et al., *Reference Frame of the Ventriloquism Aftereffect.* Journal of Neuroscience, 2009. 29(44): p. 13809-13814.

13. Kopco, N., et al., *Hemisphere-specific properties of the vnetriloquism aftereffect.* J Acoust Soc Am, 2019. 146(2): p. EL177-183.

14. Grothe, B., M. Pecka, and D. McAlpine, *Mechanisms of sound localization in mammals.* Physiol Rev 2010. 90: p. 983-1012.

15. Groh, J.M., *Making space: how the brain knows where things are.* Cambridge, MA: Harvard University Press., 2014.

16. Yao, L. and C.K. Peck, *Saccadic eye movements to visual and auditory targets.* Exp Brain Res, 1997. 115: p. 25-34.

17. Pages, D.S. and J.M. Groh, *Looking at the Ventriloquist: Visual Outcome of Eye Movements Calibrates Sound Localization.* Plos One, 2013. 8(8).

18. Gabriel, D.N., D.P. Munoz, and S.E. Boehnke, *The eccentricity effect for auditory saccadic reaction times is independent of target frequency.* Hearing Res, 2010. 262: p. 19-25.

19. Burnham, K.P. and D.R. Anderson, *Multimodel Inference Understanding AIC and BIC in Model Selection.* Sociological Methods & Research, 2004. 33(2): p. 261-304.

20. Taboga, M., *Normal distribution - Maximum Likelihood Estimation*, in *Lectures on probability theory and mathematical statistics, Third edition. Kindle Direct Publishing*.2017.

21. Carlile, S., S. Hyams, and S. Delaney, *Systematic distortions of auditory space perception following prolonged exposure to broadband noise.* Journal of the Acoustical Society of America, 2001. 110(1): p. 416-424.

22. Kopco, N., et al., *Contextual plasticity, top-down, and non-auditory factors in sound localization with a distractor*. Journal of the Acoustical Society of America, 2015. 137(EL281-287).

23. Mullette-Gillman, O.A., Y.E. Cohen, and J.M. Groh, *Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus.* Journal of Neurophysiology, 2005. 94: p. 2331-2352.

24. Porter, K.K. and J.M. Groh, *The "other" transformation required for visual-auditory integration: representational format.* Progress in Brain Research, 2006. 155: p. 313-23.

744 25. Stern, R.M. and G.D. Shear, *Lateralization and detection of low-frequency binaural*
745 *stimuli: Effects of distribution of internal delay.* Journal of the Acoustical Society of
746 America, 1996. 100(4): p. 2278-2288.

747 26. van der Heijden, K., et al., *Cortical mechanisms of spatial hearing.* Nature Reviews
748 Neuroscience, 2019. 20: p. 609-623.

749 27. Zatorre, R.J., et al., *Where is 'where' in the human auditory cortex?* Nat Neurosci,
750 2002. 5(9): p. 905-9.

751 28. Zierul, B., et al., *The role of auditory cortex in the spatial ventriloquism aftereffect.*
752 Neuroimage, 2017. 162: p. 257-268.

753 29. Michalka, S.W., et al., *Auditory spatial coding flexibly recruits anterior, but not*
754 *posterior, visuotopic parietal cortex.* Cerebral Cortex, 2016. 26(3): p. 1302-1308.

755 30. Zwiers, M.P., H. Versnel, and A.J. Van Opstal, *Involvement of monkey inferior*
756 *colliculus in spatial hearing.* J Neurosci, 2004. 24(17): p. 4145-56.

757 31. Werner-Reiss, U., et al., *Eye position affects activity in primary auditory cortex of*
758 *primates.* Current Biology, 2003. 13: p. 554-562.

759 32. Duhamel, J.-R., et al., *Spatial invariance of visual receptive fields in parietal cortex*
760 *neurons.* Nature, 1997. 389: p. 845-848.

761 33. Mullette-Gillman, O.A., Y.E. Cohen, and J.M. Groh, *Motor-related signals in the*
762 *intraparietal cortex encode locations in a hybrid, rather than eye-centered, reference*
763 *frame.* Cerebral Cortex, 2009. in press.

764 34. Wallace, M.T. and B.E. Stein, *Cross-modal synthesis in midbrain depends on input*
765 *from cortex.* Journal of Neurophysiology, 1994. 71(1): p. 429-432.

766 35. Schiller, P.H., S.D. True, and J.L. Conway, *The effects of frontal eye field and*
767 *superior colliculus ablations on eye movement.* Science, 1979. 206: p. 590-592.

768 36. Stecker, G.C., I.A. Harrington, and J.C. Middlebrooks, *Location Coding by Opponent*
769 *Neural Populations in the Auditory Cortex.* PLoS Biology, 2005. 3(3): p. e78.

770 37. McAlpine, D., D. Jiang, and A.R. Palmer, *A neural code for low-frequency sound*
771 *localization in mammals.* Nature Neuroscience, 2001. 4(4): p. 396-401.

772 38. Salminen, N.H., et al., *A population rate code of auditory space in the human cortex.* .
773 PLoS One 2009. 4:e7600.

774 39. Dingle, R.N., S.E. Hall, and D.P. Phillips, *The three-channel model of sound*
775 *localization mechanisms: interaural level differences.* J Acoust Soc Am, 2012. 131(5):
776 p. 4023-9.

777 40. Radeau, M. and P. Bertelson, *The after-effects of ventriloquism.* Quarterly Journal of
778 Experimental Psychology, 1974. 26: p. 63-71.

779 41. Radeau, M. and P. Bertelson, *The effect of a textured visual field on modality*
780 *dominance in a ventriloquism situation.* Perception and Psychophysics, 1976. 20: p.
781 227-235.

782 42. Wozny, D.R. and L. Shams, *Recalibration of Auditory Space following Milliseconds*
783 *of Cross-Modal Discrepancy.* Journal of Neuroscience, 2011. 31(12): p. 4607-4612.

784

Appendix

To examine whether auditory saccades used for responding have properties that might be important for the current modeling, responses to auditory and audiovisual stimuli in the training regions of both experiments were further analyzed (FigureA1). Two questions were addressed. First, we examined whether the observed saccades were longer or shorter depending on whether the presence of visual component/adaptation resulted in saccades that were hypometric (shorter than needed to reach the auditory target) or hypermetric (longer than needed to reach the auditory target). Such asymmetry, if observed, would suggest that some of the effects described in Section 2, e.g., the eye-centered RF effects, might have been caused by the saccade responses. Second, we evaluated whether the ratio of the magnitudes in auditory-only responses to the respective AV responses for a given AV stimulus is constant for all combinations of audiovisual stimuli. If that is the case, then, independent of any possible hypo/hypermetric dependence, the model can assume that the predicted ventriloquism aftereffect is directly related to the measured ventriloquism effect.

FigureA1A shows the biases in saccade responses from the training FP for targets in the training regions from both experiments (circles vs. squares). Open symbols represent audio-visual responses, filled symbols auditory-only responses. Black symbols represent the AV-aligned runs, while the cyan and magenta symbols represent, respectively, the runs in which the response shifts towards the visual component/adaptation resulted in saccades that were hypometric and hypermetric. Specifically, the magenta circles represent the central-training data with visual component shifted to the right, i.e., towards the fixation point, while the magenta squares represent the peripheral-training data with visual component shifted to the left, i.e., again towards the fixation point (the cyan data then represent the corresponding data for visual components shifted in the opposite direction). Note that the filled symbols here show the same data as the red lines in the training regions of Figure 1B, C.

The black symbols in FigureA1A show that, in both experiments, all the saccades in the AV-aligned runs were fairly accurate. Specifically, responses to the AV stimuli were within +/-0.5° (open black symbols) while the saccades to the auditory targets (filled black symbols) tended to be hypometric (rightward bias for targets to the left of the FP and leftward for the targets to the right) by up to 1°, except for one data point (7.5°), discussed in detail later.

Comparison of the respective magenta and cyan symbols shows that the adaptation direction (i.e., visual component displacement) that led to hypometric saccades tended to result in larger biases than the direction leading to hypermetric saccades (for example, all the magenta filled circles are clustered around the value of 3, while the corresponding cyan filled circles are around -1). To analyze this asymmetry while accounting for the biases in the AV-aligned responses, FigureA1B shows the hypometric and hypermetric data from panel A referenced to the respective baselines and plotted such that positive values always represent bias in the direction of the visual component displacement (i.e., all the cyan squares and magenta squares had their signs flipped after subtracting the baseline). The magenta open symbols show that, independent of the training region, the VE responses measured in conditions resulting in hypometric saccades were aligned with the visual component (which was separated by 5°), while the responses resulting in hypermetric saccades (open cyan symbols) only reach approximately 80% of the visual component displacement. A mixed ANOVA with a between-subject factor of Experiment (Central, Peripheral) and within-subject factors of Shift Direction (Hypometric, Hypermetric), and Azimuth (Small, Medium, Large) performed on these data confirmed these results, showing a significant main effect of shift direction ($F(1,12) = 5.78$; $p = 0.033$). The ANOVA also found a significant Azimuth X Experiment interaction ($F(2,24) = 9.71$; $p = 0.006$) reflecting a dependence of the effect on the target location that is not further considered here, and no other significant main effects or

interactions ($p > 0.1$). On the other hand, for the VAE data, no significant difference between hypometric and hypermetric saccades was observed (a similar ANOVA on these data only found a main effect of Azimuth; $F(2,24) = 7.94$; $p = 0.002$). Thus, the strong asymmetry between the hypometric and hypermetric AV data in in panel A (filled cyan vs. magenta symbols) can be ascribed to overall hypometry of the auditory saccades exhibited also by the No-Shift data (black filled symbols). Also note that there is one hypermetric AV data point for which the response referenced to baseline is near 0 (left-most filled cyan circle), not following the pattern observed for all the other points. Most likely, this inconsistency is caused by some specific characteristic of the baseline auditory-only saccades, as this point corresponds to the only black filled symbol that shows hypermetry instead of hypometry in panel A (the black filled circle at the 7.5° location).

Finally, panel C shows the observed VAE as a proportion of the observed VE (i.e., each symbol in panel C shows the ratio of the corresponding filled and open symbols from panel B). In this analysis, one subject was identified as outlier (in at least one data point the subject differed from the across-subject mean by more than 3 standard deviations). This subject is plotted separately (crosses) and not included in the across-subject graphs. For the remaining subjects, FigureA1C shows that there is a constant relationship between the induced ventriloquism effects and aftereffects such that the aftereffect is always approximately one half of the effect (with a slight tendency to grow with the target amplitude), independent of whether the shift is hypo/hypermetric or of the training region. Confirming this observation, ANOVA with the same factors as above only found a main effect of Azimuth ($F(2,22)=10.34$, $p=0.0007$). The only other factor that approached significance was Training Region ($F(1, 11)=3.83$, $p=0.076$) while all the other factors and interactions were not significant ($p > 0.15$). These results are used in the current modeling in

859 which it is assumed that there is a constant relationship between the induced ventriloquism
860 effect and aftereffect, independent of whether the induced shift is hypometric or hypermetric.
861 ---------- Insert Figure A1 around here ----------
862

Table 1: The range and increments of values of free parameters used in systematic search through the parameter space during model simulations. Ten values of each parameter were considered with either linear or quadratic spacing. Note that parameters w_E and σ_E are not used in the HC model, while all parameters are used in the HEC model.

Parameter	Range		Increments
	min	max	
h, w	0	2	linear
k	0.01	20	quadratic
c	0	1.5	quadratic
w_E	0	1	linear
σ_H, σ_E	1	20	linear

Table 2: Values of fitted model parameters and evaluation of model performance for each simulation. AICc states the criterion for a given simulation, ΔAIC is the increase in AICc for a given simulation *re*. the simulation on a given data with the minimum AICc. The underscored model names indicate the model for which there is a substantial evidence of being a better fit for the data (rounded up value of ΔAIC smaller than 2).

Simulation	Model	Fitted parameter values							Performance	
		h	k	c	w	w_E	σ_H	σ_E	AICc	ΔAIC
No Shift	HC	1.03	0.31	1.14	1.01	-	12.06	-	130.90	2.36
(Figure 3)	HEC	1.13	0.17	0.95	1.24	0.36	12.84	2.98	128.54	-
All Data	HC	0.79	0.82	1.15	0.49	-	14.21	-	444.75	7.86
(Figure 4)	HEC	0.77	0.76	1.13	0.53	0.15	13.35	4.83	436.89	-
Central	HC	1.01	5.64	0.67	0.40	-	18.79	-	176.16	5.92
(Figure 5)	HEC	0.96	5.60	0.67	0.48	0.30	18.14	5.01	170.24	-
Peripheral	HC	0.83	3.40	1.33	0.55	-	12.43	-	136.33	-
(Figure 6)	HEC	0.82	5.33	1.33	0.56	0.04	12.12	4.91	141.89	5.56

876 **Figure Captions**

877

878 **Figure 1: Experimental design and predicted and observed ventriloquism aftereffect from Kopco**
879 **et al. [12, 13]. A) Setup: nine loudspeakers were evenly distributed at azimuths from -30° to 30°. Two**
880 **fixation points were used, located 10° below the loudspeakers at +-11.25°. On training trials, audiovisual**
881 **stimuli were presented either from the central region [12] or peripheral region [13], while the subject**
882 **fixated one FP. The audiovisual stimuli consisted of a sound paired with an LED offset by -5°, 0°, or -5°**
883 **(offset direction fixed within a session). On interleaved probe trials, the sound was presented from any of**
884 **the loudspeakers while the eyes fixated either one of the FPs. B) Predicted (left-hand panels) and observed**
885 **(right-hand panels) reference frames of the ventriloquism aftereffect. Lines represent model predictions**
886 **or across-subject means of the aftereffect magnitudes for the probe trials from the AV-misaligned runs.**
887 **C) Across-subject mean aftereffect magnitudes for the probe trials from the AV-aligned runs. Note: Error**
888 **bars have been omitted for clarity. They are presented in the simulation figures in which data are**
889 **compared to models.**

890

891 **Figure 2: Structure of the HC/HEC model and illustration of its operation. A) Block diagram of**
892 **the model. The model predicts the response bias as a function of the probe stimulus location, with**
893 **additional input parameters of the fixation position, training locations, and the observed ventriloquism**
894 **effect at the training locations (rounded blocks). Two mechanisms determine the response (square blocks).**
895 **First, saccade-related bias is always present and it is not influenced by the ventriloquism signals. Second,**
896 **auditory space representation which is adapted by ventriloquism only in HC reference frame (HC model;**
897 **“HC” arrow) or in a combination of HC and EC RFs (HEC model; “HC” and “EC” arrow). Labels B, C,**
898 **D within blocks refer to respective panels below that illustrate the function of the blocks by showing the**
899 **outputs of the model components in an illustrative simulation (for training in the central region for which**
900 **the observed AV responses are nearly unbiased). B) Saccade-related bias predictions for the two fixation**
901 **points (red and blue lines). The green diamonds show the nearly zero ventriloquism effect assumed for the**
902 **predictions shown in panel C. C) Adaptation of auditory space representation resulting from the saccade-**
903 **related bias and AV response bias as shown in panel B. Diamonds represent the disparity between AV**
904 **response bias and saccade-related bias for the training FP (red), and non-training FP in HC RF (blue**

filled) and in EC RF (blue open). Lines represent predictions of auditory space adaptation induced by these disparities. D) Response bias predicted by the model as a weighted combination of biases shown in panels B and C. Values of model parameters used for the predictions of respective model components are shown along the upper frame in each panel.

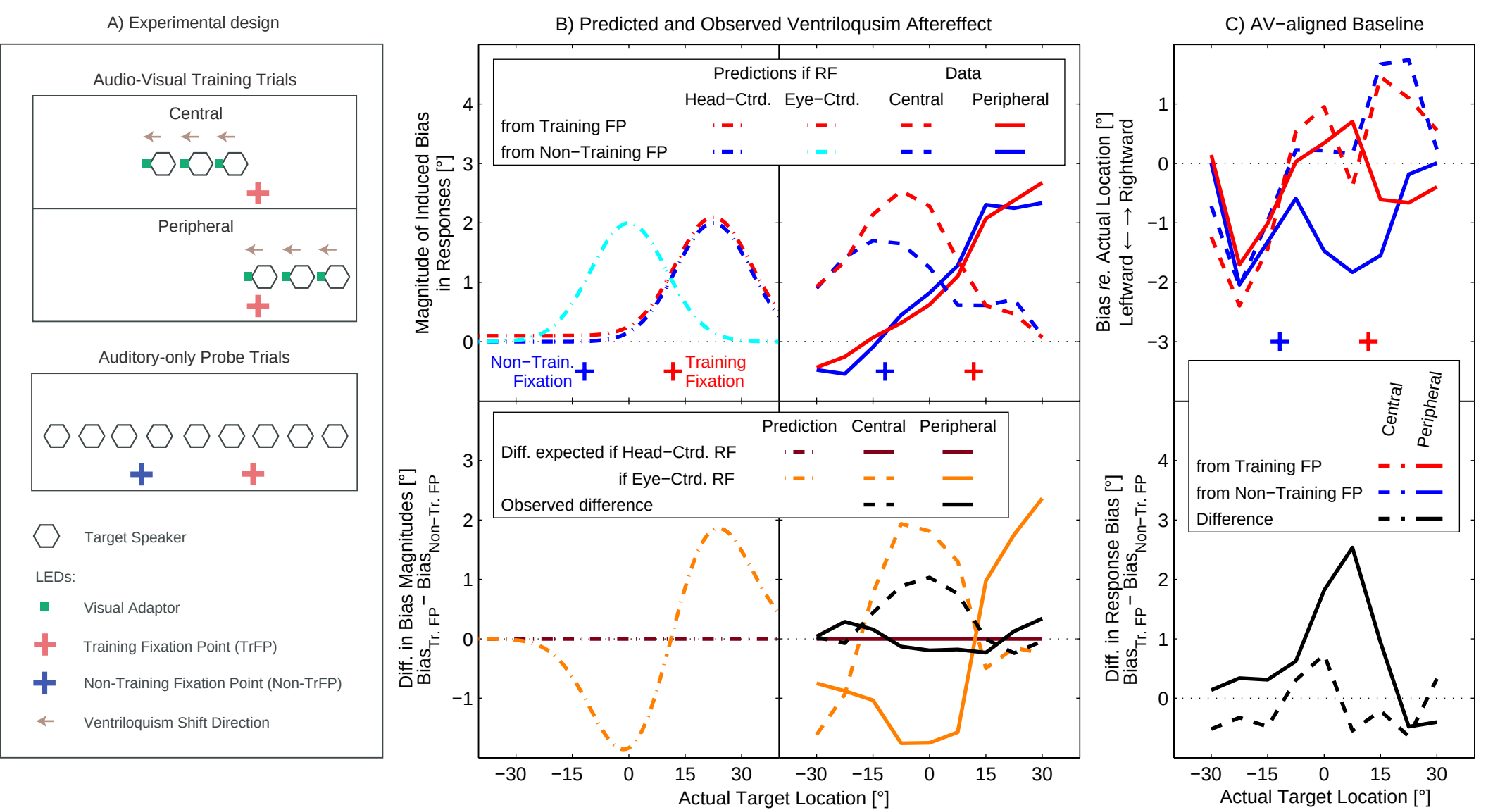
Figure 3: Model predictions and data for the No-Shift simulation. A) Visualization of the two model components, Saccade-Related Bias and Auditory Space Adaptation, for the HC and HEC models with the parameters fitted to the no-shift data (from Table 2). The Saccade-related Bias component (upper row) is independent of any visually guided ventriloquism adaptation. The Auditory Space Adaptation component (lower row) shows the strength with which the ventriloquism induced by the AV stimuli at 3 central locations shifts the responses from the Saccade-Related Bias locations to the AV response locations (Eq. 3). Note that for peripheral-training data, i.e., for the AV stimuli at the locations of 15°-30°, the lower-row graphs would be shifted by 22.5° to the right. B) Across-subject mean biases (±standard error of the mean) and model predictions for the two fixation locations (upper and middle row) and the difference between the two fixations (lower row).

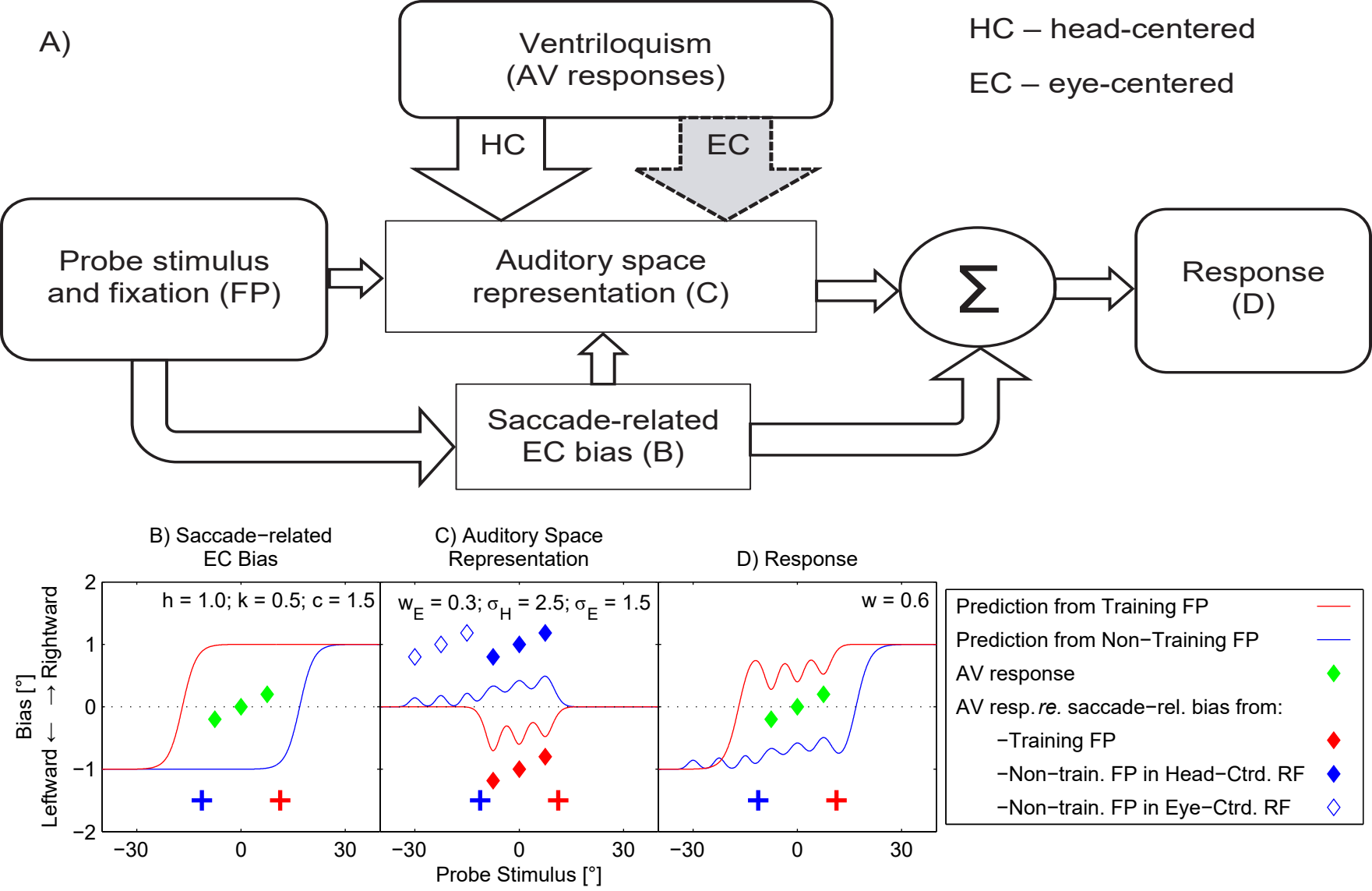
Figure 4: Model predictions and data for the All Data simulation. A) Visualization of the two model components, Saccade-Related Bias and Auditory Space Adaptation, for the HC and HEC models with the parameters fitted to all the data (from Table 2). For detailed description see the caption for panel A of Fig. 3. B) Across-subject mean difference in biases from the training FP vs. non-training FP (±standard error of the mean) and model predictions for the no-shift data, and for the aftereffect magnitude computed as a difference between positive-shift and negative-shift data (lower row).

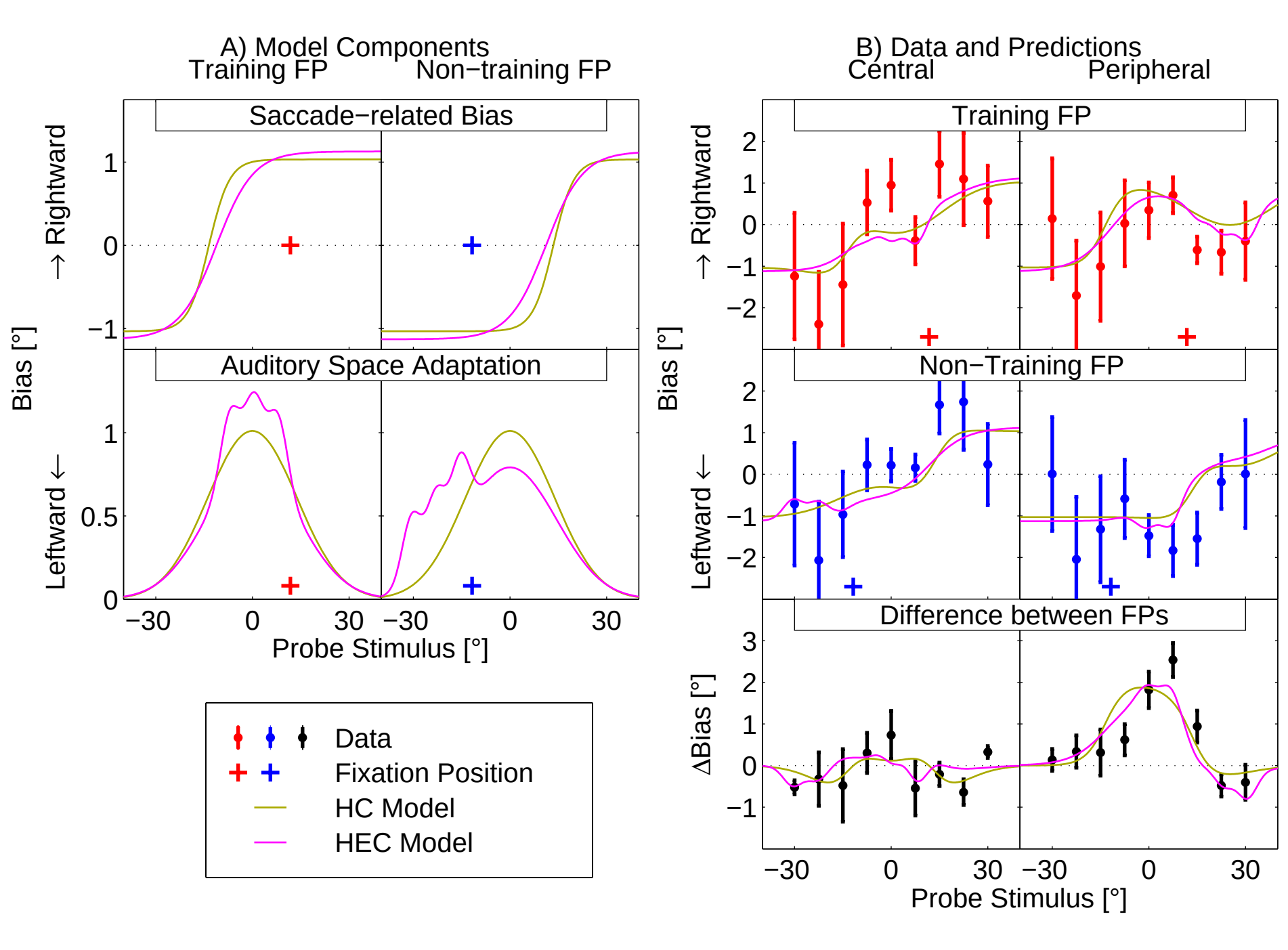
Figure 5: Model predictions and data for the Central Data simulation. For detailed description, see the caption of Figure 4.

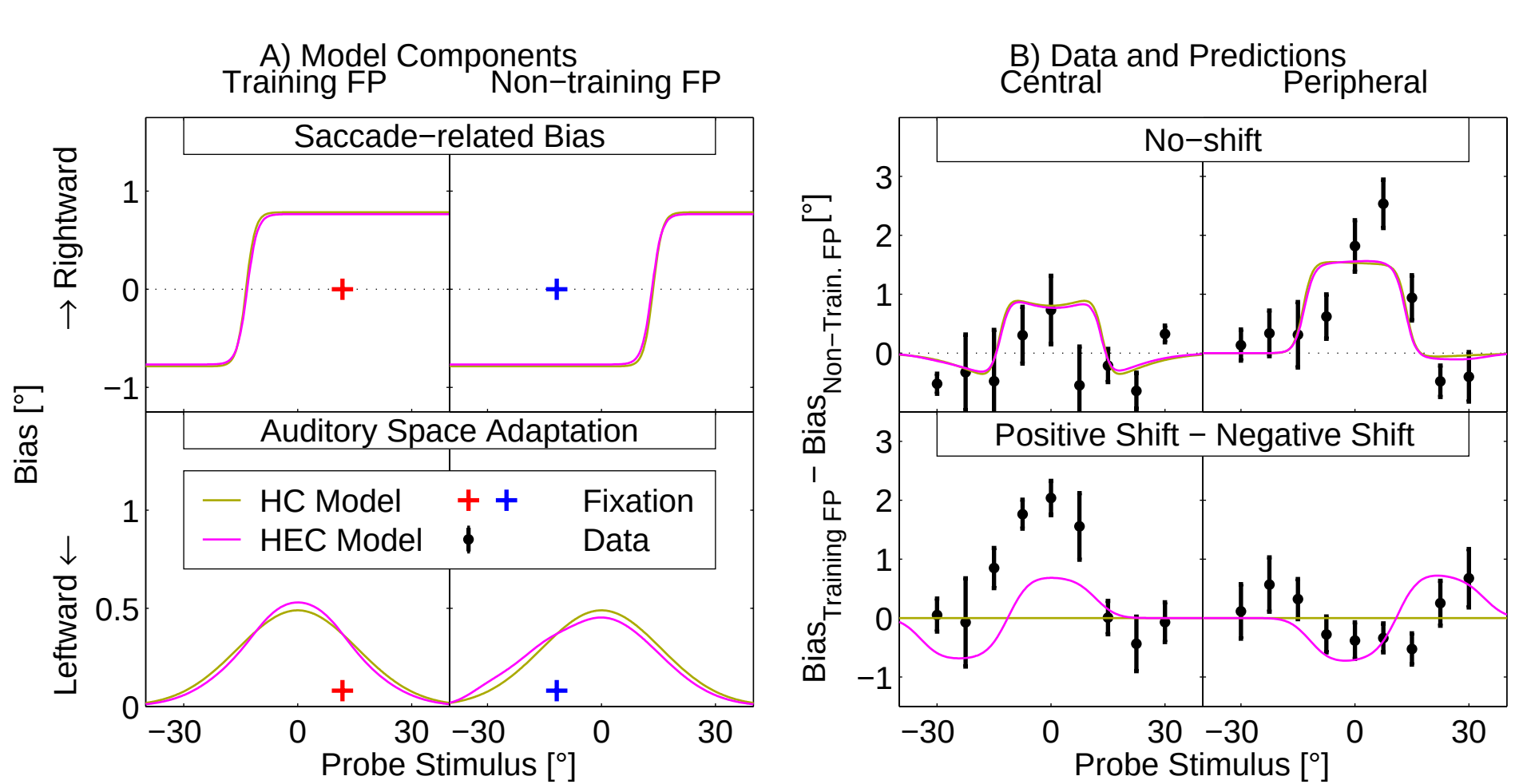
Figure 6: Model predictions and data for the Peripheral Data simulation. For detailed description, see the caption of Figure 4.

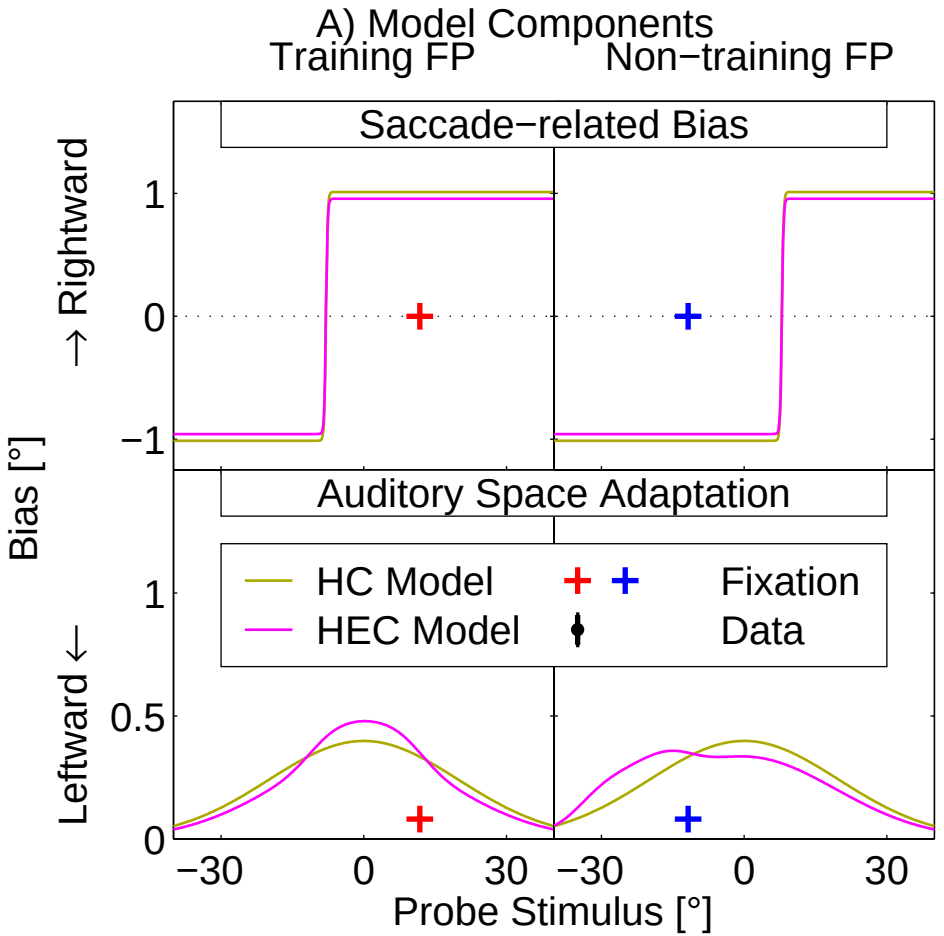
Figure A1: Saccade responses to audiovisual and auditory stimuli in the training regions from both experiments. A) Across-subject mean saccade end points as a function of the location of the auditory target or of the auditory component of the audio-visual target. Data are plotted separately for the auditory and audio-visual stimuli, for the two training regions, and for the three directions of the visual component displacement (aligned, shifting the auditory saccade to be hypometric, shifting the auditory saccade to be hypermetric). Note that a hypometric shift corresponds to visual component shifted to the right for the central-training data and to visual component shifted to the left for the peripheral training data (and vice versa for the hypermetric shift). B) Strength of the induced ventriloquism effect and aftereffect shown as the across-subject mean bias in response towards the visual component re. response in no-shift baseline (i.e., difference between the respective magenta/cyan and black symbols from panel A with the sign flipped for the negative-shift data). C) Ventriloquism aftereffect as a proportion of ventriloquism effect shown as the across-subject mean ratio of the VE/VAE strengths from panel B. Note that one outlier subject is plotted separately from the across-subject means in this analysis. Error bars represent across-subject standard errors of the means (N=7 in both experiments).

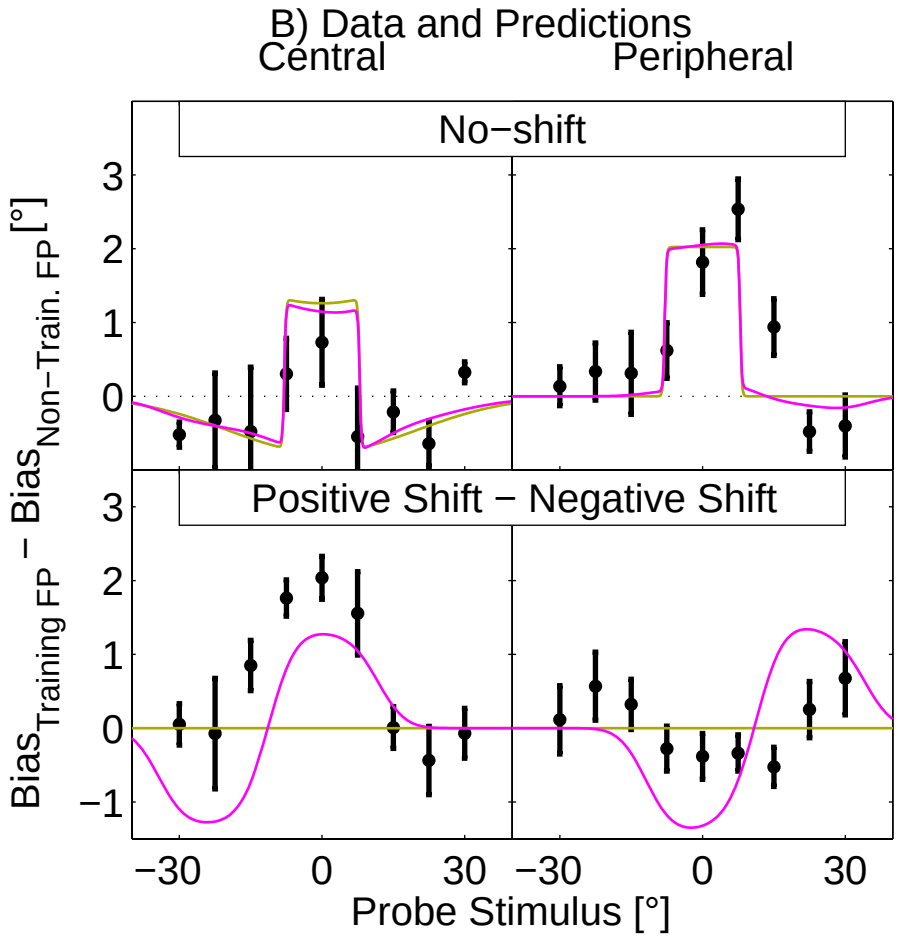

A) Experimental design
Audio-Visual Training Trials
Central
Peripheral
Auditory-only Probe Trials
Target Speaker
LEDs:
Visual Adaptor
Training Fixation Point (TrFP)
Non-Training Fixation Point (Non-TrFP)
Ventriloquism Shift Direction
B) Predicted and Observed Ventriloquism Aftereffect
Predictions if RF
Data
Head−Ctrd.
Eye−Ctrd.
Central
Peripheral
from Training FP
from Non−Training FP
Magnitude of Induced Bias in Responses [°]
Non−Train. Fixation
Training Fixation
Prediction
Central
Peripheral
Diff. expected if Head−Ctrd. RF
if Eye−Ctrd. RF
Observed difference
Diff. in Bias Magnitudes [°]
Bias Tr. FP − Bias Non−Tr. FP
Actual Target Location [°]
C) AV−aligned Baseline
Bias re. Actual Location [°]
Leftward ← → Rightward
Central
Peripheral
from Training FP
from Non−Training FP
Difference
Diff. in Response Bias [°]
Bias Tr. FP − Bias Non−Tr. FP
Actual Target Location [°]

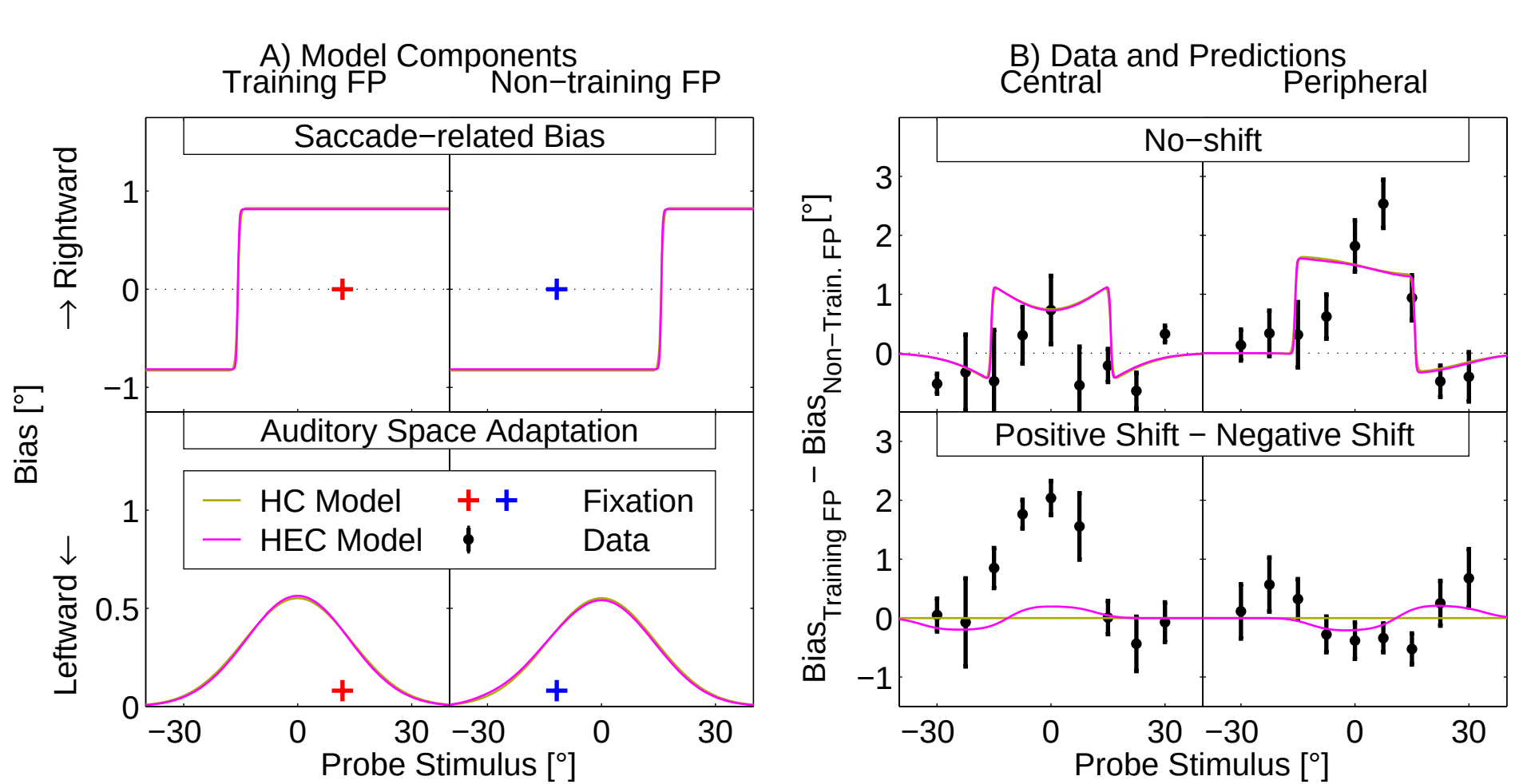

A)
HC – head-centered
EC – eye-centered
Ventriloquism
(AV responses)
HC
EC
Probe stimulus
and fixation (FP)
Auditory space
representation (C)
Σ
Response
(D)
Saccade-related
EC bias (B)
B) Saccade−related
EC Bias
C) Auditory Space
Representation
D) Response
h = 1.0; k = 0.5; c = 1.5
w_E = 0.3; σ_H = 2.5; σ_E = 1.5
w = 0.6
Bias [°]
Leftward ← → Rightward
2
1
0
−1
−2
−30
0
30
Probe Stimulus [°]
Prediction from Training FP
Prediction from Non−Training FP
AV response
AV resp. re. saccade−rel. bias from:
−Training FP
−Non−train. FP in Head−Ctrd. RF
−Non−train. FP in Eye−Ctrd. RF

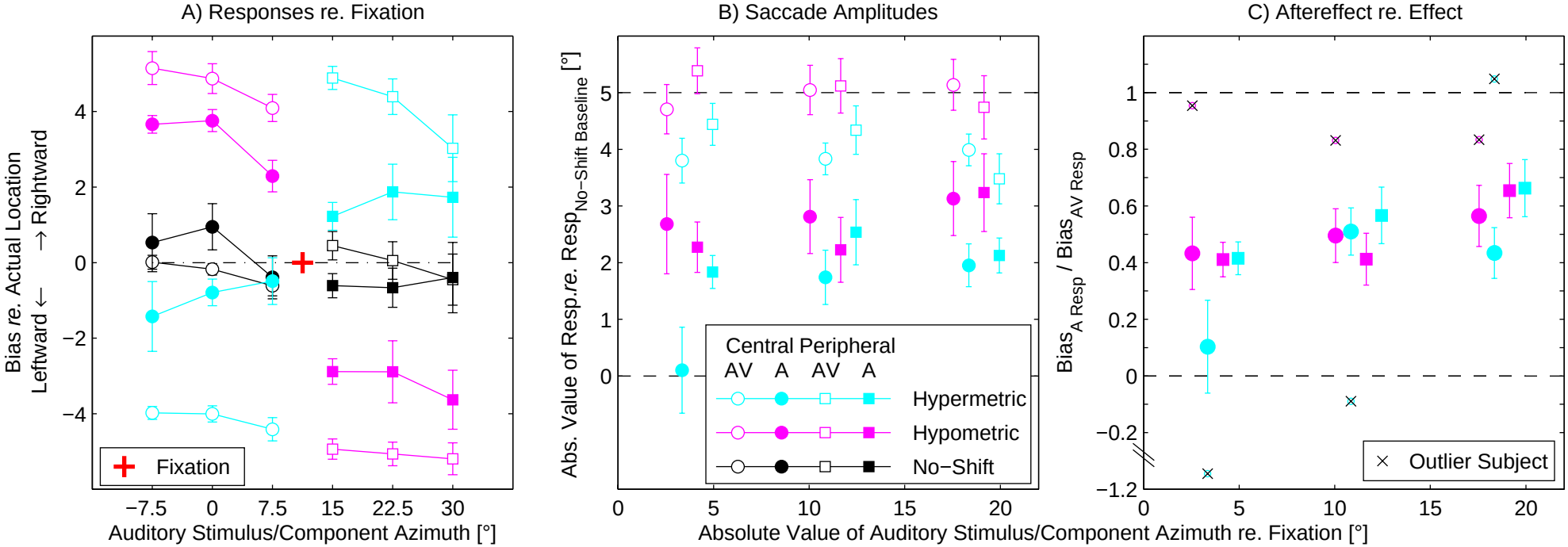

A) Model Components
Training FP
Non−training FP
Saccade−related Bias
Auditory Space Adaptation
→ Rightward
Leftward ←
Bias [°]
Probe Stimulus [°]
Data
Fixation Position
HC Model
HEC Model
B) Data and Predictions
Central
Peripheral
Training FP
Non−Training FP
Difference between FPs
ΔBias [°]


A) Model Components
Training FP
Non−training FP
Saccade−related Bias
Auditory Space Adaptation
HC Model
HEC Model
Fixation
Data
→ Rightward
Leftward ←
Bias [°]
Probe Stimulus [°]
B) Data and Predictions
Central
Peripheral
No−shift
Positive Shift − Negative Shift
Bias Non−Train. FP [°]
Bias Training FP − Bias Non−Train. FP
Probe Stimulus [°]


A) Model Components
Training FP
Non−training FP
Saccade−related Bias
Auditory Space Adaptation
HC Model
HEC Model
Fixation
Data
→ Rightward
Leftward ←
Bias [°]
Probe Stimulus [°]


B) Data and Predictions
Central
Peripheral
No−shift
Positive Shift − Negative Shift
Bias Non−Train. FP [°]
Bias Training FP − Bias Non−Train. FP [°]
Probe Stimulus [°]


A) Model Components
Training FP
Non−training FP
Saccade−related Bias
Auditory Space Adaptation
HC Model
HEC Model
Fixation
Data
→ Rightward
Leftward ←
Bias [°]
Probe Stimulus [°]
B) Data and Predictions
Central
Peripheral
No−shift
Positive Shift − Negative Shift
Bias Non−Train. FP [°]
Bias Training FP − Bias Non−Train. FP
Probe Stimulus [°]


A) Responses re. Fixation
Bias re. Actual Location
Leftward ← → Rightward
Fixation
−7.5
0
7.5
15
22.5
30
Auditory Stimulus/Component Azimuth [°]
B) Saccade Amplitudes
Abs. Value of Resp.re. Resp No−Shift Baseline [°]
Central Peripheral
AV A AV A
Hypermetric
Hypometric
No−Shift
Absolute Value of Auditory Stimulus/Component Azimuth re. Fixation [°]
C) Aftereffect re. Effect
Bias A Resp / Bias AV Resp
Outlier Subject