A Model of the Reference Frame of the Ventriloquism Aftereffect and Saccade Adaptation

Peter Lokša,[1] and Norbert Kopčo[1]

[1] *Institute of Computer Science, Safarik University, Košice, 04001, Slovak Republic*

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

Abstract: Lokša & Kopco[1] introduced a model of the reference frame of the ventriloquism aftereffect (RFoVAE) that described many but not all available RFoVAE data by assuming that the auditory spatial map, natively using the head-centered RF, is adapted by visual signals in both eye-centered and head-centered RFs. Here, the model is extended primarily by considering that, when saccade-to-sound responses are used to measure ventriloquism, the saccades also undergo an adaptation. The extended model can explain all available data, suggesting that the RFoVAE is largely head-centered, while saccadic adaptation, which is nominally eye-centered, accounts for the mixed RF observed experimentally.

1. Introduction

The "ventriloquism aftereffect" (VAE) is a cross-modal adaptation effect in which the perceived location of sounds presented alone is shifted after repeated presentations of spatially mismatched visual and auditory stimuli[2,3,4]. To induce the VAE, the visual and auditory components of the audiovisual stimuli need to be spatially transformed in the brain as their location is initially encoded using different reference frames (RFs): visual space is encoded relative to the direction of eye gaze; auditory space relative to the head orientation[5]. Existing evidence about the RF into which the signals get aligned is inconsistent. Some studies suggest the RF is mostly head-centered[6], while others suggest it is mostly eye-centered[7]. Additionally, it might be a mixture of the two RFs[8,9] , possibly reflecting the RF in which the oculo-motor responses to the stimuli are executed[10]. To examine these inconsistencies computationally, Lokša & Kopčo[1] introduced a model, called *dHEC*, that could describe some of the available data, particularly those observing a mixed RF[8,6]. However, the model could not simultaneously explain the inconsistency between the results suggesting a head-centered RF[9] and those suggesting a mixed RF[6], both of which were obtained in experiments that used saccadic eye movements to perceived auditory target location (i.e., "auditory saccades") as a response method (details of the experiments and of the model are in the following sections). The primary goal of the current study is to extend the *dHEC* model[1] to provide a unified explanation of all available data by considering that the saccades to auditory targets, used as a response method in refs. 6 and 9, also undergo adaptation during VAE training. The secondary goal is to extend the model to also describe a new adaptation

phenomenon in which shifts in responses were induced even by spatially aligned audiovisual stimuli[9] (again, described in detail in the following section). The *dHEC* model[1] proposed a mechanism that could explain the new phenomenon when considered in isolation, but not when considered simultaneously with the other ventriloquism phenomena. With these extensions, the current model aims to provide a unified theory of the visually induced adaptive phenomena in sound localization and their RF.

The current paper first summarizes the previous experimental results[6,9] and the *dHEC* model[1]. Then, the model extensions are introduced and evaluated.

2. Summary of previous experiments

Kopco et al.[6, 9] reported the results of two RF of VAE experiments that were identical in all aspects except that the first one[6] examined VAE locally induced in the central subregion of audiovisual space while the second one[9] used the peripheral subregion (Figure 1A, top panel). They used one initial eye fixation position (FP) on training trials (red '+' symbol) and presented the discrepant audiovisual stimuli from the restricted spatial range (black dashed or solid frame). On each audio-visual training trial, the auditory component (black hexagon) was presented simultaneously with the visual component (green dot) that was either aligned with it or displaced from it by 5° to the left or to the right (brown arrow; displacement direction fixed within a session). On interleaved auditory-only probe trials, they varied the initial eye position (red and blue '+' signs) with respect to the head (which was fixed) and presented sounds from locations spanning both the same

56 head-centered locations and the same eye-centered locations as on the training trials
57 (Figure 1A, bottom panel). The subject's task was to indicate the perceived location
58 of the auditory component by looking at it. No feedback was provided. The FP
59 location manipulation on probe trials served to test the RF of the recalibration, as
60 described in the following.

61 Figure 1B shows the experimental results from the AV-misaligned conditions
62 (averaged across data from sessions with visual component shifted leftwards and
63 rightwards) and expressed as bias *re.* auditory component of the AV stimuli. The
64 responses to AV training stimuli were always very near the visual components in
65 both the central and peripheral experiments (green dashed and solid lines displaced
66 by 4-5° in Figure 1B), as well as in the AV-aligned baseline (green lines in Figure 1C
67 are near 0°). The displaced V component in the AV-misaligned conditions induced a
68 local VAE when measured with the eyes fixating the training FP (the red solid and
69 dashed lines in Figure 1B show that maximum ventriloquism was always induced in
70 the trained subregion of the auditory space). The critical manipulation of these
71 experiments was that half of the probe trials were performed with eyes fixating on a
72 new, non-training FP (blue "+" symbol), shifted away from the training FP (red "+"
73 symbol). It was expected that, if the VAE is induced in a head-centered RF, then
74 moving the eyes to the new FP would have no effect (i.e., that the blue lines would
75 be aligned with the corresponding red lines). On the other hand, if it is induced in a
76 purely eye-centered RF, then the location at which the maximum VAE is observed
77 would move with the FP (i.e., the blue lines would be identical to the corresponding
78 red lines, except that they would be shifted to the left by 23.5°, i.e., the angular

separation between the FPs). The experimental data showed that, in the central experiment, moving the fixation resulted in a smaller VAE with the peak moving in the eye gaze direction (blue vs. red dashed line), while in the peripheral experiment, only a negligible effect of the FP shift was observed (blue and red solid lines overlap). To better visualize these results, the lower panel of Figure 1B shows data expressed as the difference between responses from training versus non-training FPs from the respective upper panels (each black line shows the difference between the respective red and blue data). For the central data, the black dashed line deviates from zero, showing a mixed nature of the RF of the VAE induced in this region. On the other hand, the black solid line is always near zero, corresponding to a predominantly head-centered representation. The main goal of the current study is to extend our previous model to explain this discrepancy between the central and peripheral data.

Figure 1C shows the bias induced in the baseline runs with AV training stimuli aligned. In the central-training experiment the responses from the two FPs were similar (red and blue dotted lines). In the peripheral-training experiment the responses for the targets at −7.5° to +15° differed between the two fixations, such that the non-training FP responses fell well below the training-FP responses (red vs. blue solid lines). Thus, the peripheral AV-aligned stimuli induced a fixation-dependent adaptation in the auditory-only responses in the central region, a VAE-like adaptation phenomenon that has not been previously reported. The black dashed and solid lines in Figure 1C, showing the difference between the corresponding red and blue data from the upper panel, highlight the FP-dependence

of the peripheral data in contrast to the FP-independence in the central data. The secondary goal of the current study is to extend the model to explain this inconsistency.

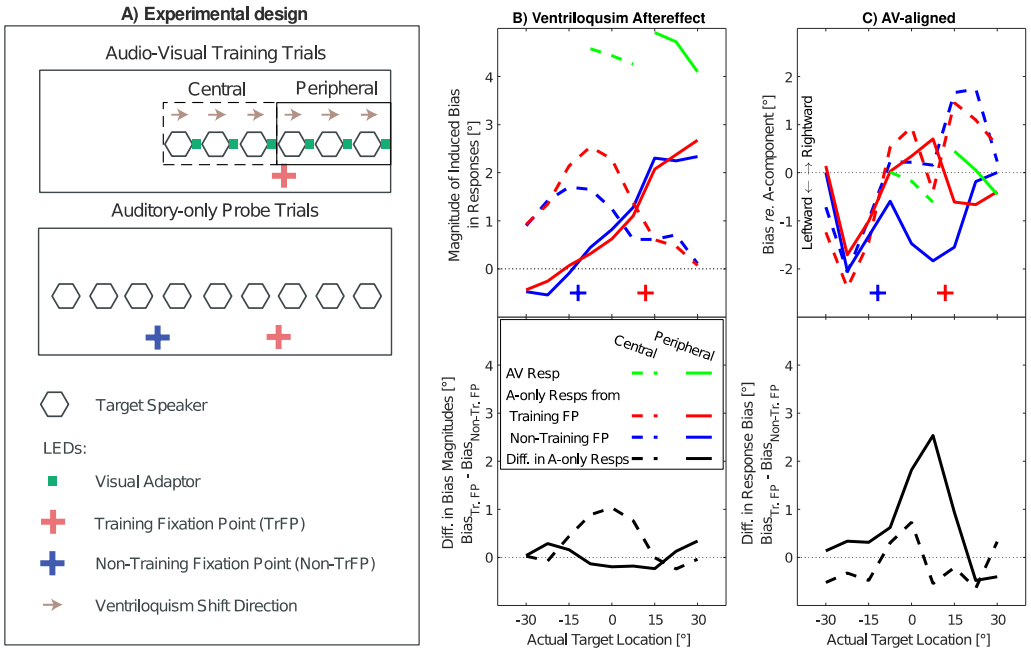

Fig. 1. Experimental design and results from Kopco et al. [6,9]. (A) Experimental design: nine loudspeakers were evenly distributed at azimuths from −30° to 30° at distance of 1.2 m. Two FPs were located 10° below the loudspeakers at ±11.75° from the center. On training trials, audio-visual stimuli were presented either from the central region (auditory-component at −7.5, 0, or +7.5° [6]) or peripheral region (auditory-component at 15, 22.5, or 30° [9]), while the subject fixated the training FP. The AV stimuli consisted of a sound paired with an LED offset by −5°, 0°, or +5° (offset direction fixed within a session). On probe trials, the sound was presented from any of the loudspeakers while the eyes fixated one of the FPs. (B) Results for

AV-misaligned training. Solid and dashed lines show measured across-subject mean biases in AV (green) and A-only trials (red and blue lines for respective FPs), corresponding, respectively, to the ventriloquism effect and aftereffect. Data from runs using V component shifted to the right and to the left are combined and always plotted as if a rightward shift was induced. Black lines show the differences between the respective red and blue lines. (C) Results for AV-aligned training plotted using the same format as in panel B. All horizontal axes are plotted in head-centered RF.

3. Model Description

Figure 2A shows the structure of the current version of the model. It predicts the VAE bias as a function of the A-only probe azimuth and the FP location (the "response bias" vs. "probe stimulus and FP" blocks). The main model component is the "auditory space representation" block that encodes the VAE biases induced by the visual ventriloquism signals ("ventriloquism" block) in head-centered coordinate frame ("HC" arrow), assuming a continuous uniform representation of auditory space. The induced VAE is determined by only considering the AV stimuli used during training. As illustrated in Fig. 2B, for each AV stimulus, the induced bias (black line) is strongest at the location of the A component of the stimulus, independent of the fixation location, and it decreases with distance from that location. Also, the model assumes that the overall strength of the bias is proportional to the measured ventriloquism effect at those locations (black line peak is below the green circle, representing the measured VAE strength).

Additional model components (gray in Fig. 2A) are optional and implement alternative mechanisms that are examined as candidates for influencing the observed RF of VAE results. The "EC" component ("EC" arrow) implements the hypothesis that the ventriloquism signals influence the auditory space representation in eye-centered RF. This mechanism was examined in the previous modeling study[1], the best-performing model from which ("*dHEC*") is used as a reference for the current model evaluation.

The main new component of the current model is the "auditory saccade adaptation" block. This block implements the hypothesis that, in addition to adaptation of the auditory spatial map, the ventriloquism also induces adaptation of the auditory saccades when saccades are used for responding to the AV and A stimuli. Specifically, for the training condition illustrated in Fig. 1A (central training region and rightward AV shift), it proposes that the saccade amplitude gets adapted to become hypometric (shorter than needed to reach the auditory target) since that agrees with the direction of the VAE-inducing AV stimulus disparity in that condition. This is illustrated in Fig. 2C, where the thick red arrow represents the rightward shift induced by the AV target (green) located to the left of the training FP (red '+' sign). This induced saccade hypometry then generalizes to new target locations from the trained FP (thin red arrows), as well as to new, non-trained FPs (blue '+' sign and the thin blue arrows), for which the hypometries result in biases that are either leftward or rightward in HC RF (the effect reverses for hypermetric adaptation, i.e., when saccade amplitude becomes longer than needed to reach the auditory target, as illustrated for the peripheral training in Fig. 1A). Additionally, the

160 effect is assumed to be only visible when the training region is between the fixation
161 regions, and its possible effects on perceived elevation are not considered. Since
162 existing studies of auditory saccades have not reported this specific form of
163 adaptation of auditory saccades[13], it is proposed here as a parsimonious mechanism
164 that can explain the available RF of VAE data.

165 The "saccade-related bias" block is an optional component introduced in
166 *dHEC* model to explain the adaptation induced by the no-shift peripheral-training
167 condition (blue, red, and black solid lines in Figure 1C). It proposes that, when
168 saccades are used as the response method, there are a priori biases in saccade
169 responses to auditory-only stimuli[11, 12] (illustrated by red and blue lines in Fig. 2D)
170 that get locally "corrected" by ventriloquism near the locations from which aligned
171 AV stimuli are presented on interleaved AV trials (blue and red arrows represent
172 shift from the a priori bias towards the AV aligned target). While in *dHEC* model this
173 mechanism could explain the newly observed adaptation when considered in
174 isolation, the predictions were less accurate when all the data of the two experiments
175 were considered simultaneously. The final new extension of the model addresses this
176 shortcoming by proposing an alternative to the simple addition of the components
177 used in the previous model. Specifically, here a scaled version of the sum of the
178 Auditory Saccade Representation and Saccade-Related Bias inputs is proposed as an
179 alternative ("Normalization" block), implementing a version of normalization to
180 limit the overall output of the neural channel[14]. The scaling (Fig. 2E) effectively acts
181 such that near the locations of the AV stimuli (green dots) the output is dominated

182 by the auditory-space-related input (green Gaussians), while outside the area it is
183 dominated by the saccade-related-bias component (black line).

184 There are three versions of the model evaluated here, all including the "Space
185 Representation" and "Saccade-related Bias" blocks and differing by whether they
186 include the optional components "Auditory Saccade Adaptation" and
187 "Normalization". The *sHC* model only includes the "Auditory Saccade Adaptation"
188 block, the *nHC* only the "Normalization" block, while the *snHC* model includes both
189 components (if "Normalization" is replaced by a simple sum, as in the previous
190 model version). They are compared to the dHEC model which performed best in the
191 previous study[1].

192 To generate the predictions, the model only uses information about the
193 training AV stimulus locations and the average measured AV response biases for
194 those locations (i.e., the green data from Fig. 1B and 1C). As described previously[1],
195 this allows the appropriate version of the model to be applied to any VAE data in
196 which the FP locations, A-component locations and AV disparities are manipulated
197 during training. Also, the model can be applied to data obtained both using the
198 auditory saccade as well as other response methods. Finally, to make the model more
199 accessible, all the model versions were integrated into the Auditory Modeling
200 Toolbox.

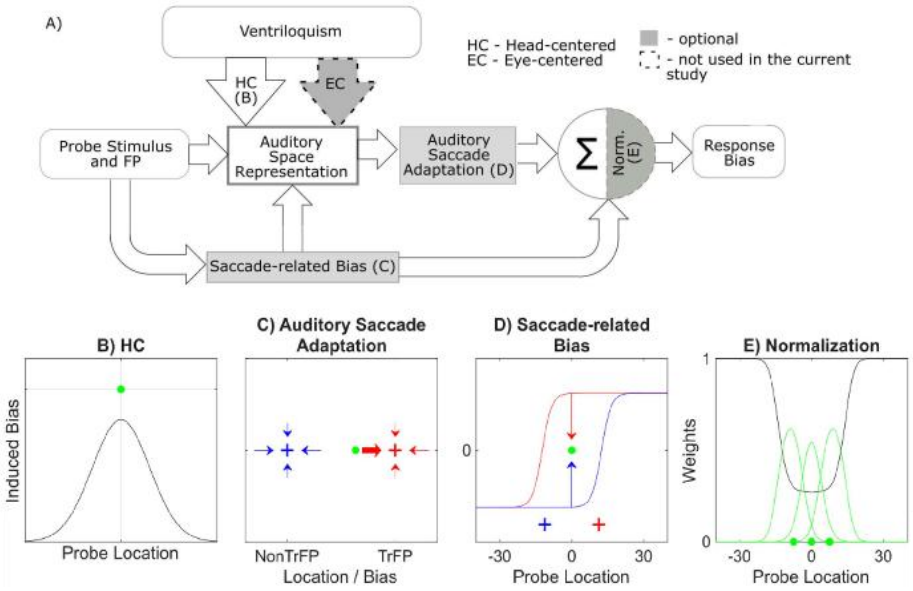

Fig. 2. Structure of the model and illustration of its operation. (A) Block diagram of the model in which the optional model components are shown in gray. Panels B through E illustrate the operation of each of the components (see text for details).

3.1 Detailed Model Specification

The following model specification applies to the most general *snHC* model version, with the differences applying to reduced versions described as needed (all variables use HC representation and are in the units of degrees).

Equation 1 describes the predicted bias in responses $\hat{r}$ to a given auditory stimulus at location s and for eyes fixating the location f as a combination of ventriloquism-induced adaptation in auditory spatial representation r_V and the

optional saccade-related bias r_E. In *dHEC*, this combination was a straightforward weighted sum:

$$\hat{r}\,(s,f) = r_E\,(s,f) + w\,r_V\,(s,f), \tag{1a}$$

which is now also normalized (as shown in Fig. 2E):

$$\hat{r}(s,f) = \frac{(1-w)\,r_E(s,f)+w\,r_V(s,f)}{(1-w)+w\,\sum_{i=1}^{N} w_{v,i}(s)}. \tag{1b}$$

Current model versions *nHC* and *snHC* use Eq. 1b, while *sHC* uses Eq. 1a. Parameter w is a free scaling parameter specifying the relative weight of the ventriloquism adaptation vs. the saccade-related bias, while $w_{v,i}$ is an additional ventriloquism-specific normalization factor defined in Eq. 3.

The ventriloquism-induced adaptation is defined as:

$$r_V(x,f) = \sum_{i=1}^{N} a_S\left(w_{v,i}(x).\left(r_{AV,i} - r_E\left(s_{AV,i},f\right)\right), s_{AV,i}, f, f_{T,i}\right) \tag{2}$$

where *i* is the index through the N combinations of training locations and FPs (in the current experiments only the training location varies and $N = 3$). For each combination, the response shift due to VAE is computed as the disparity between the ventriloquism effect responses $r_{AV,i}$ and the a priori bias $r_E\left(s_{AV,i},f\right)$ weighted by $w_{v,i}$, and then modulated by the saccade adaptation a_S. Weights $w_{v,i}$ use normal probability density functions φ, centered at the training locations with a standard deviation of σ, as a measure of the distance-dependent influence of i-th training location on target at location x in the HC frame (Fig. 2B):

$$w_{v,i}(x) = \frac{\varphi\left((x - s_{AV,i})/\sigma\right)}{\varphi(0)}. \quad (3)$$

The auditory saccade adaptation (Eq. 4) implements the mechanism illustrated in Fig. 2C. It scales the ventriloquism output by a constant amount, g, such that its effect is either in the same direction or in the opposite direction of the VAE depending on whether current FP f is at the same side of the current AV training location s_{AV} as it was during the training f_T (for the *nHC* model version, $g = 0$):

$$a_S(x, s_{AV}, f, f_T) = x \,.\left(1 - g\,\psi(m(s_{AV} - f_R))\right), \quad (4)$$

$$\text{where } f_R = f \text{ if } \left(f - s_{AV,i}\right)\left(f_T - s_{AV,i}\right) < 0, \text{and } f_R = f_T \text{ otherwise.}$$

The ψ represents a sigmoidal function (Eq. 6) with a slope determined by parameter m. The condition for f_R proposes that the saccade adaptation occurs with respect to the current fixation only if the training region was between the current and training fixations (like in the current "central training" data), while it remains at the training fixation otherwise (e.g., in the current "peripheral training" data). The behavior of this function is further described in the Results section.

Finally, the saccade-related bias represents the hypothesized *a priori* shifts in responses (red and blue curves in Fig. 2D) as a sigmoid centered at a fixed distance from the fixation f:

$$r_E(x, f) = h\,\psi\left(k(x + cf)\right), \quad (5)$$

where

$$\psi(x) = \frac{2}{1+\exp(-x)} - 1\,. \quad (6)$$

To produce predictions, the model uses information about the locations of the training AV stimuli $s_{AV,i}$, training FPs $f_{T,i}$, and the measured ventriloquism biases for those stimuli $r_{AV,i}$, obtained from the experimental data (here, green lines in Fig. 1) and it fits the free model parameters *w,* σ*, g, m, h, k,* and *c*.

3.2 Parameter fitting and model evaluation

All evaluation procedures, briefly summarized here, were similar to ref. 1. The complete data set used in the simulations consists of the AV-aligned and AV-misaligned data for the central and peripheral training regions shown in Figure 1B–C. The training FP and non-training FP data, as well as their difference were used (blue, red, and black lines in Figure 1), resulting in a data set contained 108 A-only across-subject mean and standard deviation stimulus-response data points ([9 azimuths] × [2 FPs +FP difference] × [2 AV conditions] × [2 training regions]). The corresponding AV training stimuli and responses (green lines in Figure 1) were used as fixed model parameters.

The three models were fitted to the data using a two-step procedure that minimized the weighted MSE, consisting of a systematic search through the parameter space followed by a non-linear iterative least-squares fitting (Matlab function lsqnonlin). To compare the models' performance while accounting for the number of parameters used by each model, we computed the Akaike information

criterion AICc. We use the rule that the model with the lower AICc is substantially better than an alternative model only if the rounded-up value ΔAIC is larger than 2. Also, the mean absolute error (MAE) was evaluated for each model (and used for AICc computation).

4. Results

Figure 3 presents the simulation results by showing the experimental data (from Fig. 1, now with SEM error bars) and the fitted models (lines), separately for the training FP (top row), non-training FP (middle row), and their difference (bottom row). Each column shows the results for a different combination of training region (central vs. peripheral) and AV stimuli (aligned vs. misaligned). For the experimental data, the first notable observation is that the error bars on the FP-difference plots (black data in the bottom panel) are much smaller than those for the individual FPs (red and blue in top and central panels). Therefore, the critical evaluation of the current models was performed on the FP-difference data (bottom row), while the overall trends in the data for each FP considered separately (top and middle rows) are largely captured by all the models. The FP-difference plots for the AV-misaligned data can be interpreted as showing the extent to which there is an EC contribution to the RF in the data, with values near zero representing a purely HC RF, while the larger the deviations from zero, the larger the contribution of EC RF. The *dHEC* model[1] performs poorly, as it predicts a similar amount of the EC and HC RF contributions for the central and peripheral data (gray dashed lines in the bottom rows in Fig. 3A and 3B have similar peaks), while the data either have a larger peak (Fig. 3A) or no peak at all (Fig. 3B). While the *nHC* model cannot predict

this pattern of results at all (beige line is fixed at zero in both panels), the models with the saccade adaptation component can predict it very accurately (magenta and green lines in both panels).

Comparison of the *snHC* and *nHC* model predictions in the upper and middle panels of Fig. 3A illustrates the function of the saccade-adaptation mechanism for the central-training data. Since this region lies on opposite sides re. the two FPs, saccade adaptation acts in the opposite direction for the two FPs, increasing the adaptation effect for the training FP (magenta and green lines are above beige line in the upper panel) while decreasing it for the non-training FP (magenta and green lines are below beige line in the lower panel). On the contrary, in peripheral training the saccade adaptation always decreases the adaptation strength (green and magenta lines are below the beige line in both the top and the middle panel of Fig. 3B), thus inducing no difference between the FPs.

For the AV-aligned data, the *dHEC* model predicts a similar amount of FP-disparity for the central and peripheral experiments (gray dash-dotted line in the bottom panels of Figs. 3C and 3D), while the data show a much larger disparity in the latter experiment. While the current *sHC* model predictions are similarly inaccurate (magenta line is near gray line in both Figs. 3C and 3D), the normalization component of the *nHC* and *snHC* models mostly produces a much closer match to the data (beige and green lines in the two panels). The upper and middle panels show that the normalization mechanism allows the models to follow the data more closely for both fixations and both experiments (i.e., green lines are more often closer to the

data than magenta lines), however the match is hard to evaluate given how noisy the data are (large error bars on the red and blue data).

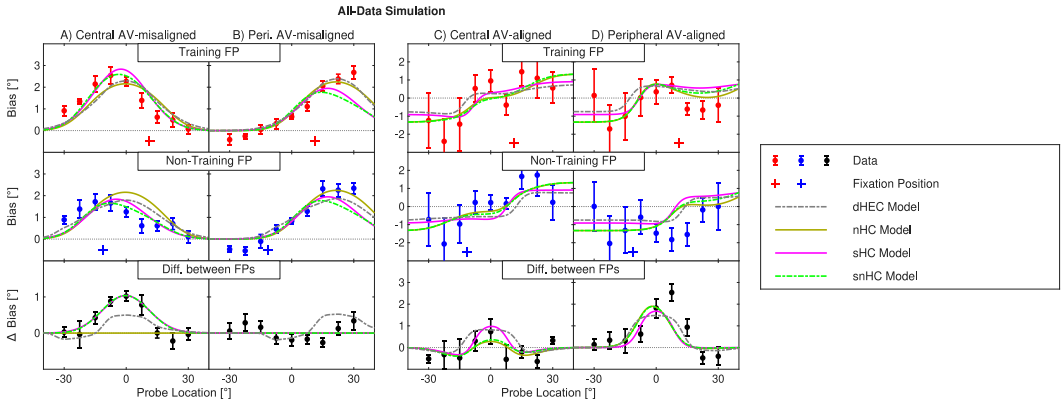

Fig. 3. Model predictions (lines) and experimental data (symbols). Each column corresponds to one combination of central vs. peripheral and AV-aligned vs. AV-misaligned data. Top and middle rows: Across-subject mean biases (±SEM) and model predictions for the two FPs separately. Bottom row: Differences between the biases (±SEM) for the two individual FPs and corresponding differences between the model predictions from the top and middle rows.

Table 1. Fitted model parameters and model performance for the current models and the dHEC model[1]. ΔAIC is the increase in AICc for a given model re. the model with the lowest AICc. The underlined model names indicate the model version with substantial evidence of better fit to the data (i.e., rounded up ΔAIC larger than 2).

Model version	Performance			Parameters						
	AICc	ΔAIC	MAE	h	k	c	w	σ	g	m

nHC	345.6	6.9	1.30	1.33	0.30	0.74	0.27	11.15	-	-
sHC	340.8	2.1	1.19	0.91	0.40	0.70	0.19	12.41	0.50	0.09
snHC	338.7	-	1.17	1.33	0.29	0.72	0.24	11.12	0.45	0.11
dHEC	352.7	14.0	1.31	See Lokša & Kopčo[1].						

331

332

333 The performance of the models and the fitted parameter values are
334 summarized in Table 1. The MAE values confirm that the *snHC* model is the most
335 accurate, and the AICc shows that this improvement is considerable (ΔAIC > 2), in
336 particular with respect to the previous *dHEC* model. However, a comparison of the
337 AICc values for the three model versions also shows that it is mainly the saccade
338 adaptation mechanism component that improved the overall model performance, as
339 the omission of this component causes a much larger increase in the AICc than the
340 omission of normalization.

341 **5. Discussion**

342 The current study introduced a model of the reference frame of the
343 ventriloquism aftereffect by extending a previous model[1] to consider adaptation in
344 auditory saccades and a normalized combination of model component outputs. This
345 model can predict all available human data, which the *dHEC* model[1] could already
346 predict[8], as well as the Kopco et al. data[6,9] which were successfully predicted only
347 with the current model extensions. The challenge in providing a unified description
348 of the Kopco et al. data lies in that the two studies provide conflicting results, the

former suggesting that the RF of VAE is mixed, while the latter one only observing a HC RF. The current model proposes that the RF of VAE is purely HC, while the mixed-RF result of ref. 6 was driven by the adaptation in auditory saccades which were used as a response method in both those studies. Additionally, the current version of the model can simultaneously predict the new ventriloquism-like adaptation induced by AV-aligned stimuli[9]. For that, it proposes that the effect of the "saccade-related bias" mechanism is combined with the ventriloquism-induced biases using a normalized summation instead of simple summation proposed previously.

A primary assumption of this model is that the auditory spatial representation is uniform, consisting of narrow spatial channels. There is growing evidence that, in mammals, auditory space is primarily encoded based on two or more spatial channels roughly aligned with the left and right hemifields of the horizontal plane[15]. Considering such opponent processing might provide an alternative mechanism for predicting the differences between the central and peripheral data considered here (see Ref. 1).

The mechanisms of auditory saccade adaptation and saccade-related bias proposed in this model need to be experimentally tested, as there is overall a scarcity of data about the auditory saccades and their adaptation[13,11]. Also, it needs to be examined whether these are two different mechanisms, as implemented here, or whether they are actually a part of one system controlling the auditory saccades[10]. Finally, the mechanisms proposed here can be linked with the existing physiological

371 evidence about RF of the signals in the auditory pathway[7] and about adaptation and
372 normalization of auditory spatial representation[14].

373 In summary, the current model provides testable predictions for future RF of
374 VAE studies using both auditory saccades and other response methods. Its
375 implementation containing all components introduced here and in ref. 1 is publicly
376 available and incorporated into the Auditory Modeling Toolbox to enhance its future
377 testing on new data or generation of predictions for new experiments.

378 **DATA AVAILABILITY**

379 A Matlab / Octave implementation of the model and the experimental data
380 are available in the Auditory Modeling Toolbox
381 (https://sourceforge.net/p/amtoolbox/code/ci/loksa2025/tree/models/loksa2025.
382 m).

383 **ACKNOWLEDGMENTS**

384 This research was supported by APVV-23–0054, APVV SK-AT-23-0002,
385 and HORIZON-MSCA-2022-SE-01 Grant No. 101129903.

386 **Declaration of Conflicting Interests**

387 The authors declared no potential conflicts of interest with respect to the
388 research, authorship, and/or publication of this article.

389 **References and Links**

[1]P. Lokša, and N. Kopčo, "Toward a Unified Theory of the Reference Frame of the Ventriloquism Aftereffect," *Trends in Hearing*. **27** (2023), doi: 10.1177/23312165231201020.

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

[3]G. H. Recanzone, "Rapidly induced auditory plasticity: The ventriloquism aftereffect," Proc Natl Acad Sci U S A **95**, 869–875 (1998). doi:DOI 10.1073/pnas.95.3.869

[4]T. M. Woods, and G. H. Recanzone, "Visually Induced Plasticity of Auditory Spatial Perception in Macaques," Current Biology **14**, 1559–1564 (2004).

[5]J. M. Groh, and D. L. Sparks, "Two models for transforming auditory signals from head-centered to eye-centered coordinates," Biol Cybern **67**, 291–302 (1992).

[6]N. Kopco, I. F. Lin, B. G. Shinn-Cunningham, and J. M. Groh, "Reference Frame of the Ventriloquism Aftereffect," Journal of Neuroscience **29**, 13809–13814 (2009). doi:Doi 10.1523/Jneurosci.2783-09.2009

[7]V. C. Caruso, D. S. Pages, M. A. Sommer, and J. M. Groh, "Compensating for a shifting world: Evolving reference frames of visual and auditory signals across three multimodal brain areas," J Neurophysiol **126**, 82–94 (2021). doi:10.1152/jn.00385.2020

[8]D. M. Watson, M. A. Akeroyd, N. W. Roachi, and B. S. Webb, "Multiple spatial reference frames underpin perceptual recalibration to audio-visual discrepancies," PLoS ONE **16** (2021) doi:10.1371/journal.pone.0251827

[9]N. Kopco, P. Loksa, I. F. Lin, J. Groh, and B. Shinn-Cunningham, "Hemisphere-specific properties of the ventriloquism aftereffect," J Acoust Soc Am **146**, EL177-183 (2019).

[10]J. van Opstal, *The Auditory System and Human Sound-Localization Behavior*, (Elsevier, 2016), 1st ed. doi:https://doi.org/10.1016/C2014-0-00203-1

[11]L. Yao, and C. K. Peck, "Saccadic eye movements to visual and auditory targets," Exp Brain Res **115**, 25–34 (1997).

[12]D. N. Gabriel, D. P. Munoz, and S. E. Boehnke, "The eccentricity effect for auditory saccadic reaction times is independent of target frequency," Hearing Res **262**, 19–25 (2010).

[13]J. Hu, and P. Vetter, "How the eyes respond to sounds.," Ann N Y Acad Sci. Feb;1532 (2024) doi:10.1111/nyas.15093

[14]J. C. Dahmen, P. Keating, F. R. Nodal, A. L. Schulz, and A. J. King, "Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space," Neuron **66**, 937–948 (2010). doi:DOI 10.1016/j.neuron.2010.05.018

[15]B. Grothe, M. Pecka, and D. McAlpine, "Mechanisms of sound localization in mammals.," Physiol Rev **90**, 983–1012 (2010).

430