Modeling the mixed reference frame of the ventriloquism aftereffect

Peter Lokša & Norbert Kopčo

Institute of Computer Science, P. J. Šafárik University, Košice, Slovakia

1. ABSTRACT

Ventriloquism aftereffect (VA) is observed as a shift in the perceived locations of auditory stimuli, induced by repeated presentation of audiovisual signals with incongruent locations of auditory and visual components. Since the two modalities use a different reference frame (RF), audition is head-centered (HC) while vision is eye-centered (EC), the representations must be aligned before the visual re-calibration can occur. Previous studies examining RF of VA found inconsistent results: the RF was observed to be a mixture of HC and EC for VA induced in the center of the audiovisual field [Kopčo et al., J.Neurosci.29, 13809–13814, 2009], while it was predominantly HC for VA induced in the periphery [Kopčo et al., JASA 146, EL177, 2019]. In addition, the latter study found an adaptation in the auditory space representation even for congruent AV stimuli. Here, a computational model examines the origins of these effects, focusing on the question whether the neural visual signals guiding spatial auditory plasticity use both HC and EC RFs, or only the HC RF.

Two versions of the model are evaluated, both consisting of two main processing stages that interact additively: (1) a stage of auditory spatial representation (using HC RF) which receives the visual calibration signals, and (2) a stage of saccadic eye response control (using EC RF) which introduces a priori biases even in no-shift baseline responses and which can correct the a priori biases when responding to aligned audiovisual stimuli. The two versions of the model differ in whether it is assumed that the 1st stage processes visual calibration signals exclusively in HC coordinates (HC model) or in both HC and EC coordinates (HEC model). Four simulations were performed, evaluating the models' predictions for different aspects of the data. The critical simulation evaluated the two model versions on the complete data sets from the two previous studies, to determine whether the presence of EC-referenced visual signals of the HEC model significantly improves the HEC model performance compared to the HC model. Here Akaike information criterion AICc is used to evaluate the models.

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

These results suggest that visual signals in both HC and EC reference frames are used to calibrate the naturally HC auditory spatial representation, even when EC-referenced eye-saccade adaptation is accounted for. However, there still were important aspects of data that neither of the two model versions could describe, suggesting that non-linear adaptive processes also contribute to the observed results.

2. BACKGROUND AND INTRODUCTION

- Several previous models were developed to describe the ventriloquism aftereffect in humans and birds. There are models of the audio-visual (AV) RF alignment, but those only consider AV integration (Razavi et al., 2007) and multi-sensory integration (Pouget et al., 2002) when in the auditory and the visual stimuli are presented simultaneously (i.e., the ventriloquism effect; VE), not the adaptation and transformations underlying VAE.
- 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 in reference frames (RFs) of auditory and visual spatial encoding.
- We primarily examine the RF in which VAE occurs. The main modeling goal is to determine 1) whether a uniform, location-independent spatial adaptation mechanism can explain the location-dependent results, and whether 2) it is only driven by head-orientation referenced visual signals, or whether signals in eye-centered RF also contribute.
- The second goal is to separate the effect of auditory saccade adaptation from the modeled RF of the VAE.
- Finally, Kopco et al. (2019) observed a new adaptive phenomenon induced by aligned audiovisual stimuli presented in the periphery. The current model proposes a mechanism of a priori biases in the responses, possibly due to auditory saccade adaptation, that can describe this phenomenon.

3. EXPERIMENTAL DATA OF KOPCO ET AL. (2009, 2019)

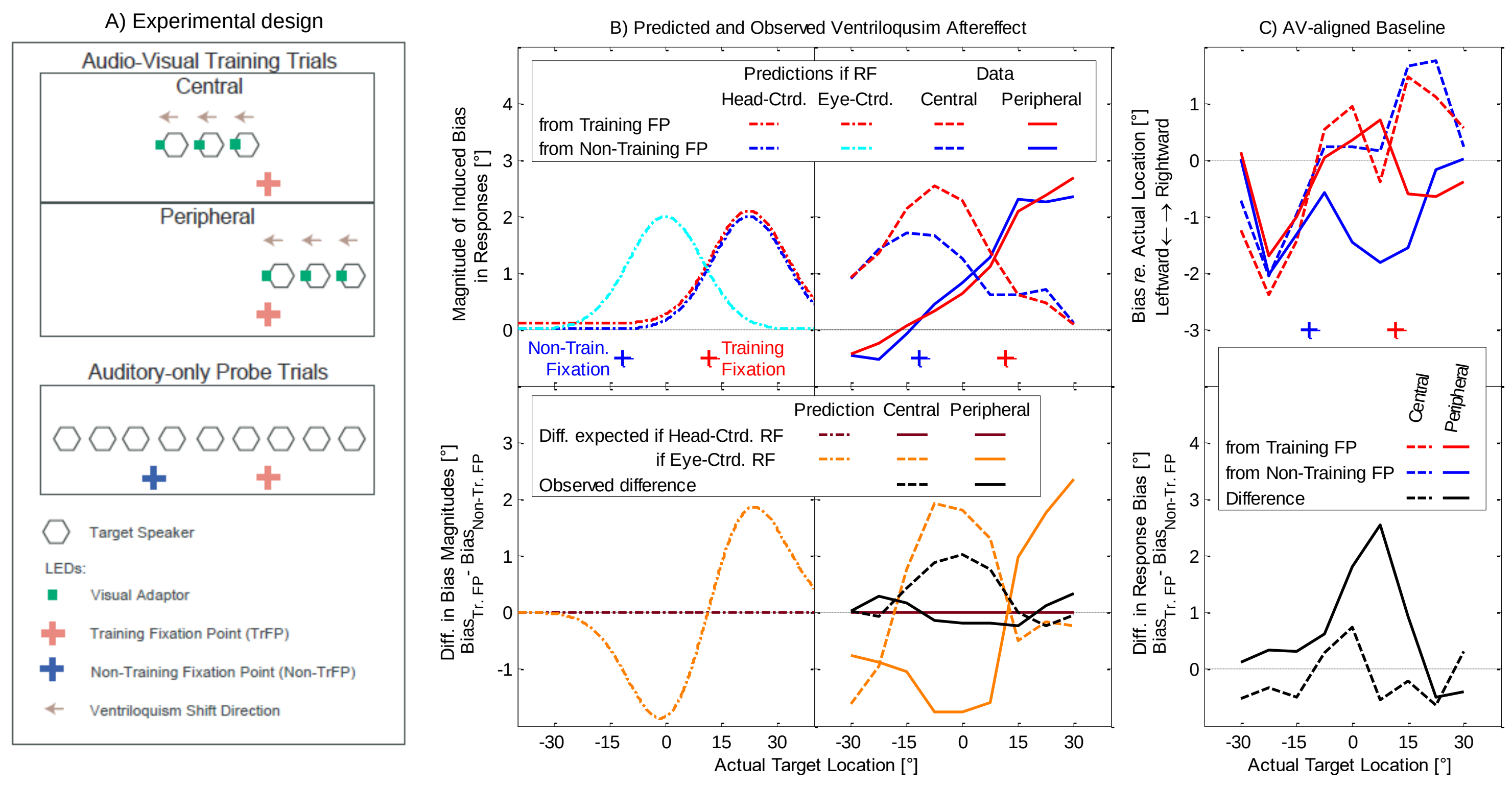


Fig. 1. A) Experimental stimuli and setup from Kopčo et. al. (2009, 2019). B) Experimental results for conditions with visual components shifted re. auditory components. C) Localization bias for no-shift AV-aligned baseline condition.

Methods and predictions

- VAE induced with eye-gaze fixed at one fixation point (FP), called training FP, using AV stimuli with V-component shift direction fixed within session (Fig. 1A). Two expreiments, each examining RF in a different training region: central, peripheral.
- VAE measured from two different FPs: training and non-training FP.
- If induced response bias **shifts with FP** then RF is **eye-centered**; if response bias **does not shift with** FP, then RF is **head-centered** (Fig. 1B).

Setup and stimuli:

- A stimuli: 300ms broadband nosie, V stimuli: LEDs synchronized with sound.
- AV stimulus disparity (fixed within session): no shift (0°); positive shift (V offset 5° to the right of A); negative shift (V shifted 5° to the left of A).
- VE and VAE responses: saccades from FP to the perceived location of auditory component.
- Trials with A-only stimuli (50%) and AV stimuli (50%) interleaved.
- AV stimuli presented with eyes fixated at training FP.
- A-only stimuli presented with eyes fixated on training or non-training FP.

Positive & Negative Shift Results (Fig. 1B):

- for central training region: RF is **mix** of head- and eye-centered
- for peripheral training region: almost **purely head-centered**,
- thus **inconsistent results** for different training regions.

No-Shift Results (Fig. 1C):

- Central training: responses independent of FP (blue and red lines overlap),
- Peripheral training: responses depend on FP (red line above blue line for central region),
- **Unexpected form of plasticity** observed for central locations with peripheral training.

Modeling questions:

- **Can adaptation in saccadic responses (as opposed to EC representation of auditory space) explain the differences in the AV-aligned baseline data?**
- **Can this adaptation also explain the differences in RFs based on the AV-misaligned data, or is it necessary that EC component is also considered, and thus that the RF is indeed mixed?**

4. AUDITORY SACCADES AND VENTRILOQUISM AFTEREFFECT

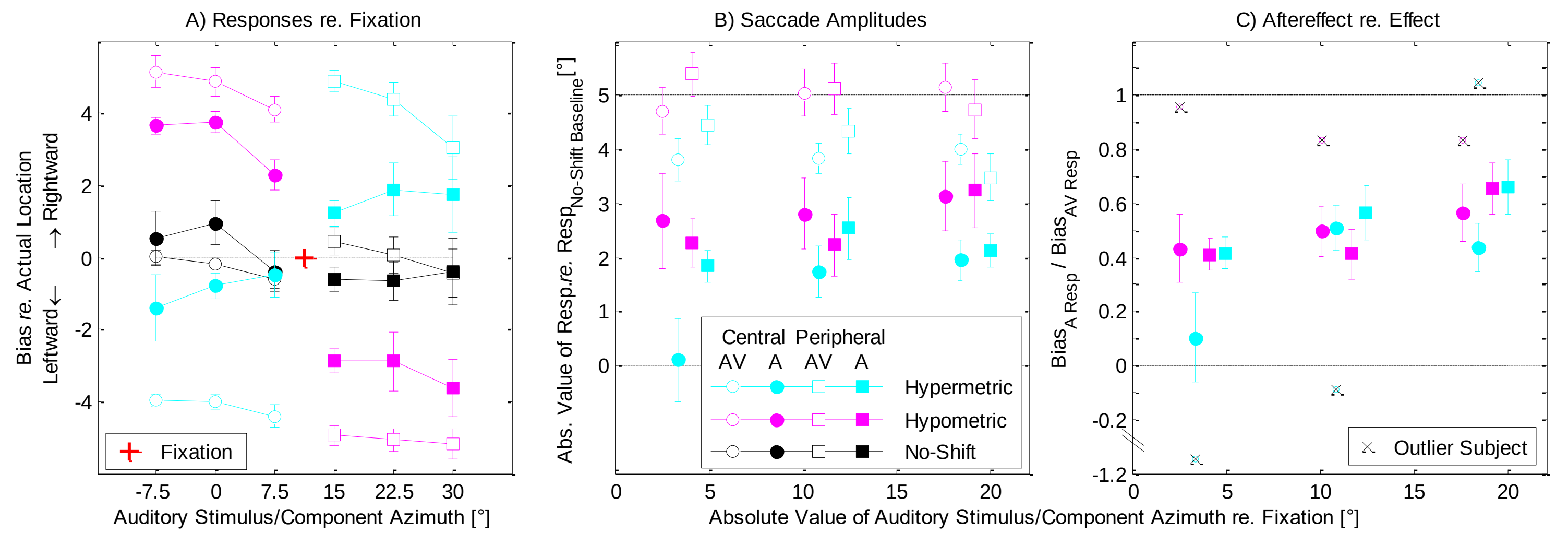


Fig. 2: A) Bias in raw saccade responses from Training FP for VE and VAE sessions that result in hypometric vs. hypermetric adaptation. Only responses for training region, where AV stimuli were presented, are further considered. B, C) Relative strength of VE and VAE hypometric and hypermetric adaptation as a function of desired amplitude (i.e., distance from FP to A-target). B) AV and A-only data from Fig. 4B scaled by the physical AV disparity. C) VAE as a proportion of VE from panel A.

Kopco et al. (2009, 19) studies were performed such that eye saccades were used both to induce ventriloquism and to evaluate VE and VAE. Saccades use eye-centered RF. Therefore, if asymmetrical VAE is induced for conditions resulting in hypometric vs. hypermetric saccades, that might contribute to the observed eye-centered component of the RF of VAE. **Here, re-analysis of the Kopco et al. (2009, 19) data to test hypo/hypermetric effects and relation between VE and VAE strength.**

Fig. 2A shows that, when raw biases are considered, the hypometric (eg. right-ward ventriloquism shift for a saccade that goes to the left) VE/VAE appears much **stronger** than hypermetric VE/VAE (respective cyan vs magenta symbols).
Fig. 2B shows that the asymmetry is weaker for VAE when referenced to no-shift baseline. But, it is still present in the ventriloquism effect, such that hypometric VE is approx. 100%, while hypermetric VE is around 80% of V-offset size.
Fig. 2C VAE expressed as a proportion of VE is approximately 50%, independent of hypo/hyper. A 3-way mixed ANOVA with between-subject factor of Training Region (central vs. peripheral) and within-subject factors of Direction (hypometric vs. hypermetric) and Amplitude (small, medium, large) only found a main effect of Amplitude (F(2,22)=10.34, p=0.0007).

Hypo/hyper asymmetry in saccades visible mainly for VE. For modeling purposes, VAE can be considered as independent of saccade shift direction, as long as VE saccades are used as reference.

5. MODEL DESCRIPTION

Computational model (Fig. 3) predicted bias for an A-only target (from a fixed FP and for a given set of AV responses) is a weighted sum (determined by weight *w*) of:

- **Saccade-related EC bias** independent of the visual signals, caused, e.g., by hypometry of saccades, inherent bias toward the periphery,
- **Bias caused by adaptation to visual signals**, defined as proportional shift towards the AV-responses, dependent on distance of the A-only target from each AV-response. This bias is independent of properties of auditory saccades (section 4)

Two versions of the model examined:

- **HC:** ventriloquism signals converted to HC reference frame for adaptation,
- **HEC:** ventriloquism signals in both HC and EC RFs adapt auditory representation.

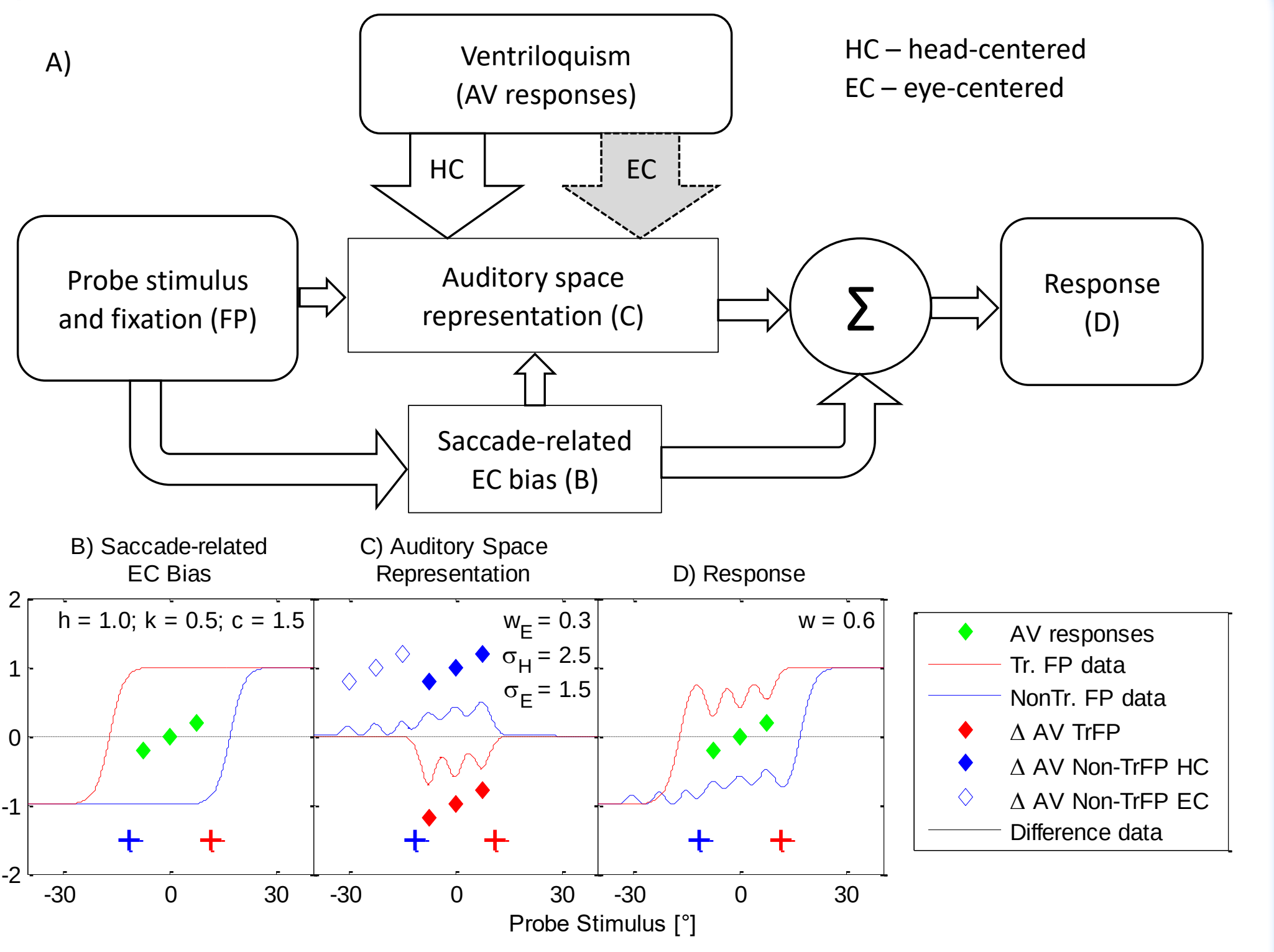


Fig. 3: A) **Model Diagram.** Response to auditory stimuli is predicted, depending on the FP location, as a linear combination of **saccade-related EC bias** (B) and **adaptation in auditory space representation** induced by the ventriloquism and proportional to the VE in the AV responses (C). The ventriloquism signals are either exclusively in HC RF (model version HC) or in a mixture of HC and EC RFs (model HEC), represented by additional gray EC branch. **B)** Saccade-related EC-bias is modeled as a sigmoid with height *h*, slope *k*, and center-offset *c*, only dependent on the FP location and the current A-stimulus location. Here the bias is shown for the 2 FP locations. **C)** VAE is induced in auditory spatial representation by AV stimuli and is proportional to observed VE and to the distance of the current A stimulus from the training AV locations either in HC RF (independent of FP location; filled symbols; HC model) or as a weighted sum of HC and EC RF signals (filled and open blue symbols). The dependence on distance is Gaussian with widths of σ_H for HC RF r σ_E for EC RF. Parameter w_E defines relative weight of the two coordinates. **D)** The effects of the two model components are summed up to produce a prediction of the response. Here, parameter *w* is the weight of the Ventriloquism adaptation (C) re. the saccade-related bias (B).

6. MODELING RESULTS

Four simulations were performed (Tab. 1 shows fitted model parameters and AICc used to compare the models):

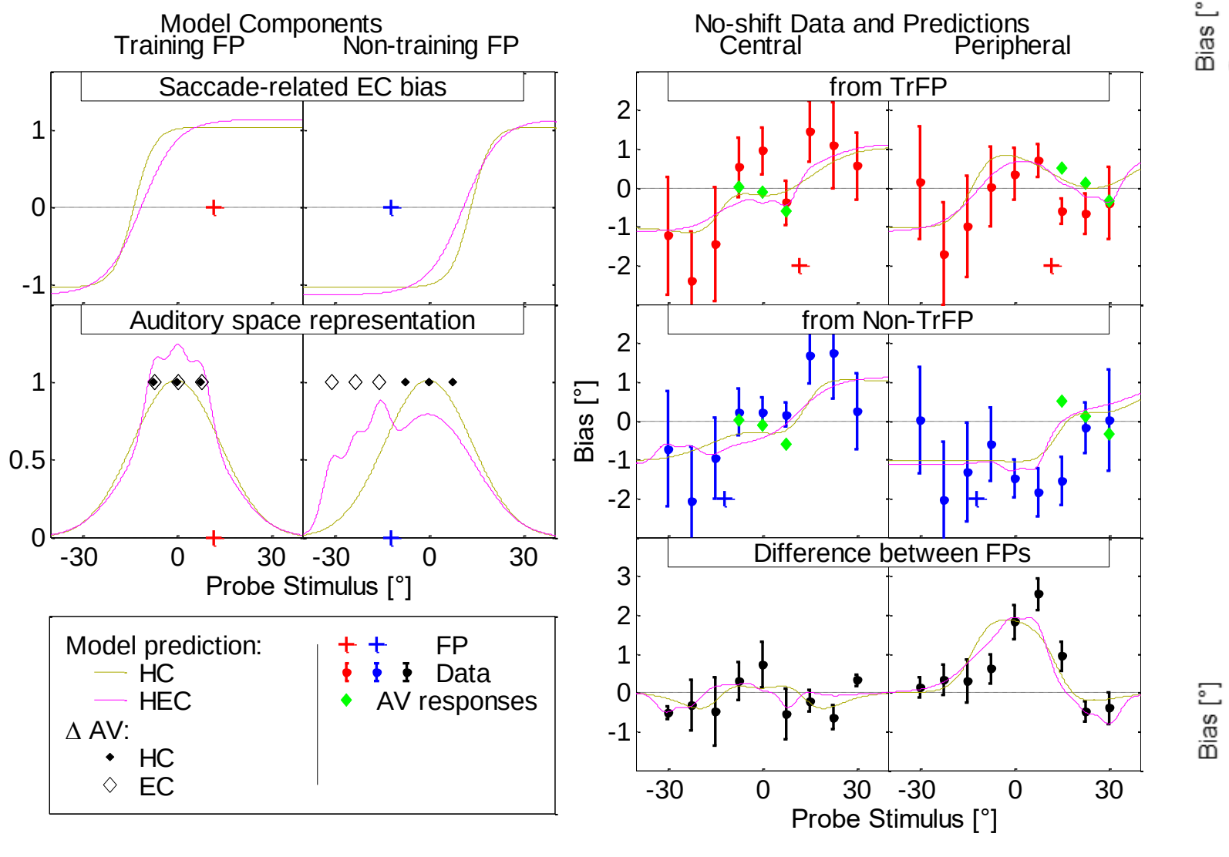
No-shift data (Fig. 4A): Both models equally good at describing data → **Adaptation in saccadic responses can explain FP-dependence of no shift data.**

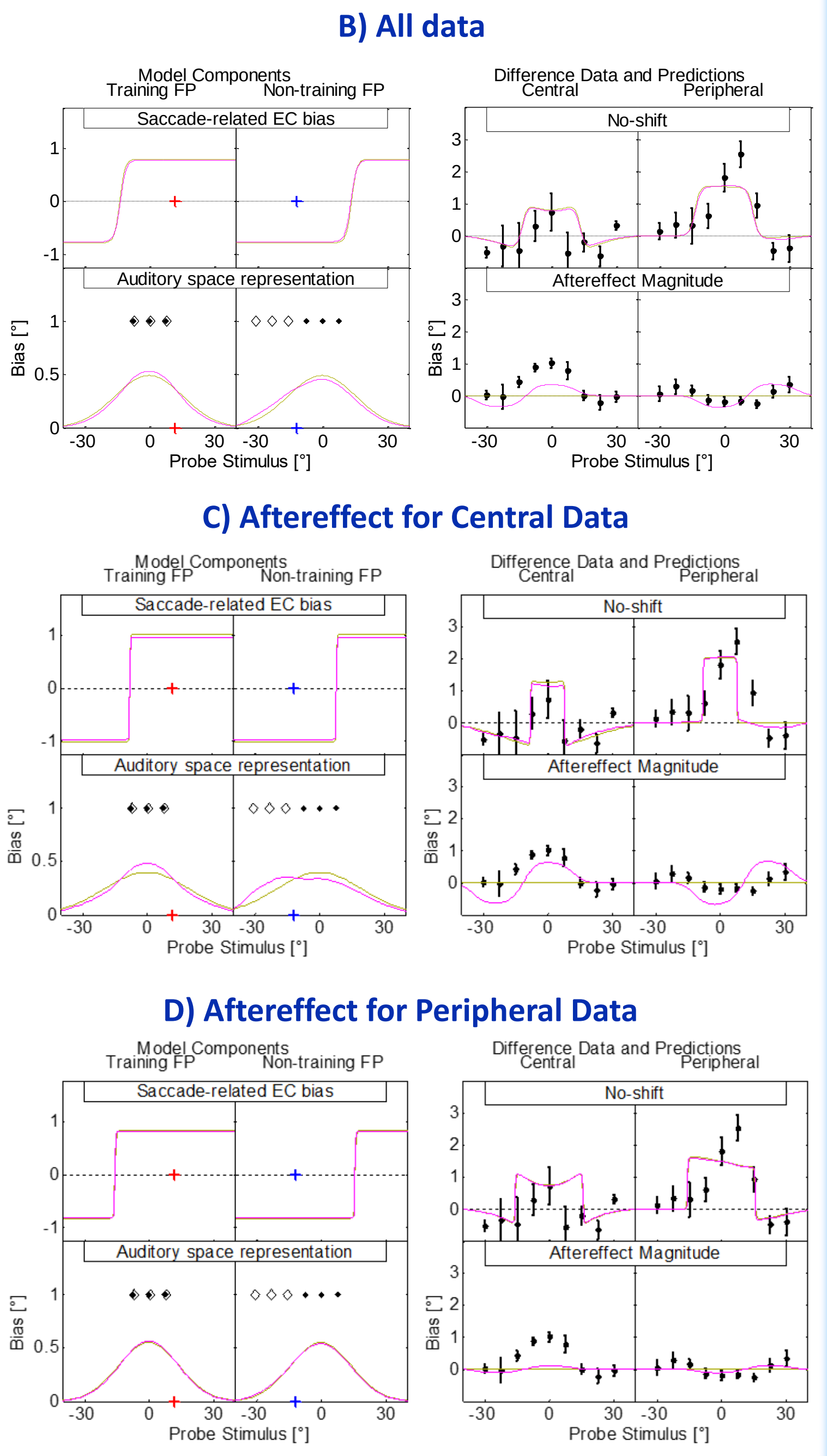
All data (shift and no-shift data for center and periphery Fig. 4B): HEC considerably better than HC → **uniform representation/adaptation of auditory space requires both HC & EC signals. the EC contribution to the auditory space representation is required** in addition to the adaptation in saccadic responses**.**

Two additional simulations were performed, fitting the model only to the **central-training aftereffect data (Fig. 4C)** and to the **peripheral-training aftereffect data (Fig. 4D)**. Central-training data require the EC component (HEC model better) while the peripheral-training data do not (HC model better).

Parameters:
- w_E always less than 0.5 → EC RF contribution weaker than HC.
- σ_E always narrower than σ_H. → EC RF signals more specific than HC.
- *w* between ~0.5 and 1 → both model components important.

A) No-shift Data





B) All data

C) Aftereffect for Central Data

D) Aftereffect for Peripheral Data

Figure 4: Performance of the models fitted in 4 simulations. Left-hand portion of each panels shows for both models the two model components (saccade-related and ventriloquism bias). Right-hand portion shows the data and the predictions of the two models. **A)** Simulation only considering no-shift data. The right-hand portion shows (from top to bottom) predictions for the training-FP data, non-training FP data, and their difference. **B)** Simulation considering all the data. Here, the right-hand portion only shows the difference between the FP data for the no shift data (top) and for the aftereffect magnitude (bottom). **C)** Simulation with models only fitted on the central aftereffect data. Right-hand portion is the same as in panel B. **D)** Simulation with models only fitted on the peripheral aftereffect data. Right-hand portion is the same as in panel B. Error bars represent across-subject standard errors of mean (N = 7).

Simulation	Model	Fitted parameter values							Performance	
		h	k	c	w	w_E	σ_H	σ_E	AICc	DiffAIC
No Shift (Fig. 4A)	Model HC	1.0322	0.3134	1.1419	1.0101	-	12.0635	-	130.9	2.4
	Model HEC	1.1284	0.1750	0.9452	1.2438	0.3636	12.8375	2.9816	128.5	0
All Conditions (Fig. 4B)	Model HC	0.7850	0.8231	1.1515	0.4893	-	14.2145	-	444.7	7.9
	Model HEC (+)	0.7659	0.7571	1.1251	0.5296	0.1458	13.3535	4.8285	436.9	0
Central Aftereffect (Fig. 4C)	Model HC	1.0115	5.6372	0.6720	0.3979	-	18.7883	-	176.2	5.9
	Model HEC (+)	0.9571	5.5968	0.6719	0.4784	0.3014	18.1408	5.0090	170.2	0
Peripheral Aftereffect (Fig. 4D)	**Model HC (+)**	0.8275	3.4037	1.3337	0.5511	-	12.4334	-	136.3	0
	Model HEC	0.8156	5.3329	1.3340	0.5631	0.0394	12.1228	4.9067	141.9	5.6

Table 1: Resulting parameter values and AICc related evaluation. **(+)** indicates model with substantial support (DiffAIC > 2.5)

7. CONCLUSIONS AND DISCUSSION

- We proposed a model of saccade responses to auditory targets after ventriloquism adaptation to describe the reference frame of ventriloquism aftereffect data of Kopco et al. (2009, 2019).
- The HC version of the model can predict the newly reported adaptation by AV-aligned stimuli (Kopco et al., 2019) as a combination of saccade-related biases "corrected" by visually guided adaptation.
- However, the HC model cannot sufficiently describe the differences between reference frames observed in central training (Kopco et al., 2009) vs. peripheral training (Kopco et al., 2019). Instead, a model that assumes that EC-referenced signals adapt the auditory representation (HEC model) is required → **Uniform reference frame of VAE is mixed, using both HC and EC referenced signals**.
- Ventriloquism adaptation considered here is local. → It is inconsistent with the models based on the opponent processing channels.
- Ventriloquism effect measured by auditory saccades is stronger if resulting in hypometric saccades (vs. hypermetric). Ventriloquism Aftereffect does not show this asymmetry. Considering VAE as a proportion of VE eliminates it almost completely.
- Future steps: experimentally test the prediction that saccade-related EC bias occurs and influences RF of VAE measurements.

ACKNOWLEDGMENTS

Work supported by Slovak Science Grant Agency VEGA-1/0355/20

REFERENCES

Alais, D. and Burr, D., (2004). The ventriloquist effect results from near-optimal bimodal integration. Curr Biol. 14:257-262
Knudsen, E. I. and Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. Science. 230:545-548
Knudsen, E. I. and Knudsen, P. F. (1989) Vision calibrates sound localization in developing barn owls. Journal of Neuroscience. 9:3306-3313
Kopčo, N., Lin, I., Shinn-Cunningham, B. G. and Groh, J. M., (2009). Reference Frame of the Ventriloquism Aftereffect. Journal of Neuroscience. 29(44): 13809-13814
Kopčo, N., Lokša, P., Lin, I., Groh, J. M., Shinn-Cunningham, B. G. (2019). Hemisphere-specific properties of the ventriloquism aftereffect. JASA 146, EL177
Maier, J. K., McAlpine, D., Klump, G. M., Pressnitzer, D., (2009). Context Effect in the Discriminability of Spatial Cues. JARO 11(2): 319-328
Brainard, M. S., Knudsen, E. I. (1995) Dynamics of visually guided auditory plasticity in the optic tectum of the barn owl. J Neurophysiol 73:595–614
Pouget, A., Deneve, S., Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. Nat Rev Neurosci. 2002 Sep;3(9):741-7. doi: 10.1038/nrn914
Razavi, B., O'Neill, W.E., Paige, G.D. (2007) Auditory spatial perception dynamically realigns with changing eye position. J Neurosci 27:10249–10258
Harris, C. M. (1994), Does saccadic undershoot minimize saccadic flight-time? A Monte-Carlo study. Vision Res. 35(5): 691-701
Lokša, P. and Kopčo, N., (2016). Modelling of the Reference Frame of the Ventriloquism Aftereffect. Sborník z 16.ročníku konference Kognice a Umělý život. 101-106
Yao, LJ and Peck, CK, (1997). Saccadic eye movements to visual and auditory targets. Springer: vol. 115; 25-34