A model of the reference frame of the ventriloquism aftereffect considering saccade adaptation

Peter Lokša & Norbert Kopčo
Institute of Computer Science, P. J. Šafárik University, Košice, Slovakia

1. ABSTRACT

Background: The ventriloquism aftereffect (VAE), observed as a shift in the perceived locations of sounds after audio-visual stimulation, requires reference frame (RF) alignment since hearing and vision encode space in different frames (head-centered vs. eye-centered). Previous experimental studies observed inconsistent results: a mixture of head-centered and eye-centered frames for the VAE induced in the central region vs. a predominantly head-centered frame for the VAE induced in the periphery. A previous model proposed to describe these data required different parameter fits to predict the central vs. peripheral data. Here, a new version of the model is introduced to provide a unified prediction of both data sets considering that saccade responses used to measure VAE are also adapted.
Methods: The model has two components: an auditory space representation component and a saccade-representation component. The former is adapted by ventriloquism signals in the head-centered reference frame. The later one characterizes adaptation in auditory saccade responses in eye-centered frame.
Results: The updated version of the model provides a unified prediction for both central and peripheral aftereffect data, even if only head-centered RF is considered in the auditory space representation.
Conclusion: The results suggest that purely head-centered RF is used for adaptation of auditory spatial representation in the ventriloquism aftereffect, and that the apparently mixed eye-and-head centered RF observed experimentally is most probably due to saccade-related biases that are eye-centered. However, additional simulations need to be performed to determine whether eye-centered ventriloquism signals further improve the model predictions.

2. BACKGROUND AND INTRODUCTION

- Existing models of the audio-visual (AV) RF alignment only consider integration when in the auditory and visual stimuli are presented simultaneously (i.e., the ventriloquism effect; VE) (Razavi et al., 2007; Pouget et al., 2002).
- We proposed a model of the visually guided adaptation of auditory spatial representation in VAE (Lokša & Kopčo, 2021) to describe behavioral data of Kopčo et al. (2009, 2019). Here extensions of the model are introduced:
  - to characterize the mixed RF of VAE observed in Kopčo et al. (2009)
  - to provide a unified account of conflicting results of Kopčo et al. (2009, 2019)
- In addition to auditory space representation in HC RF, the models consider 3 candidate mechanisms underlying these effects:
  - eye-centered signals influencing auditory space representation,
  - fixation-position-dependent attenuation in auditory space adaptation,
  - adaptation in the saccades used for responding in the experiments.
- Finally, Kopco et al. (2019) observed a new adaptive phenomenon induced by aligned audiovisual stimuli presented in the periphery is also considered.

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

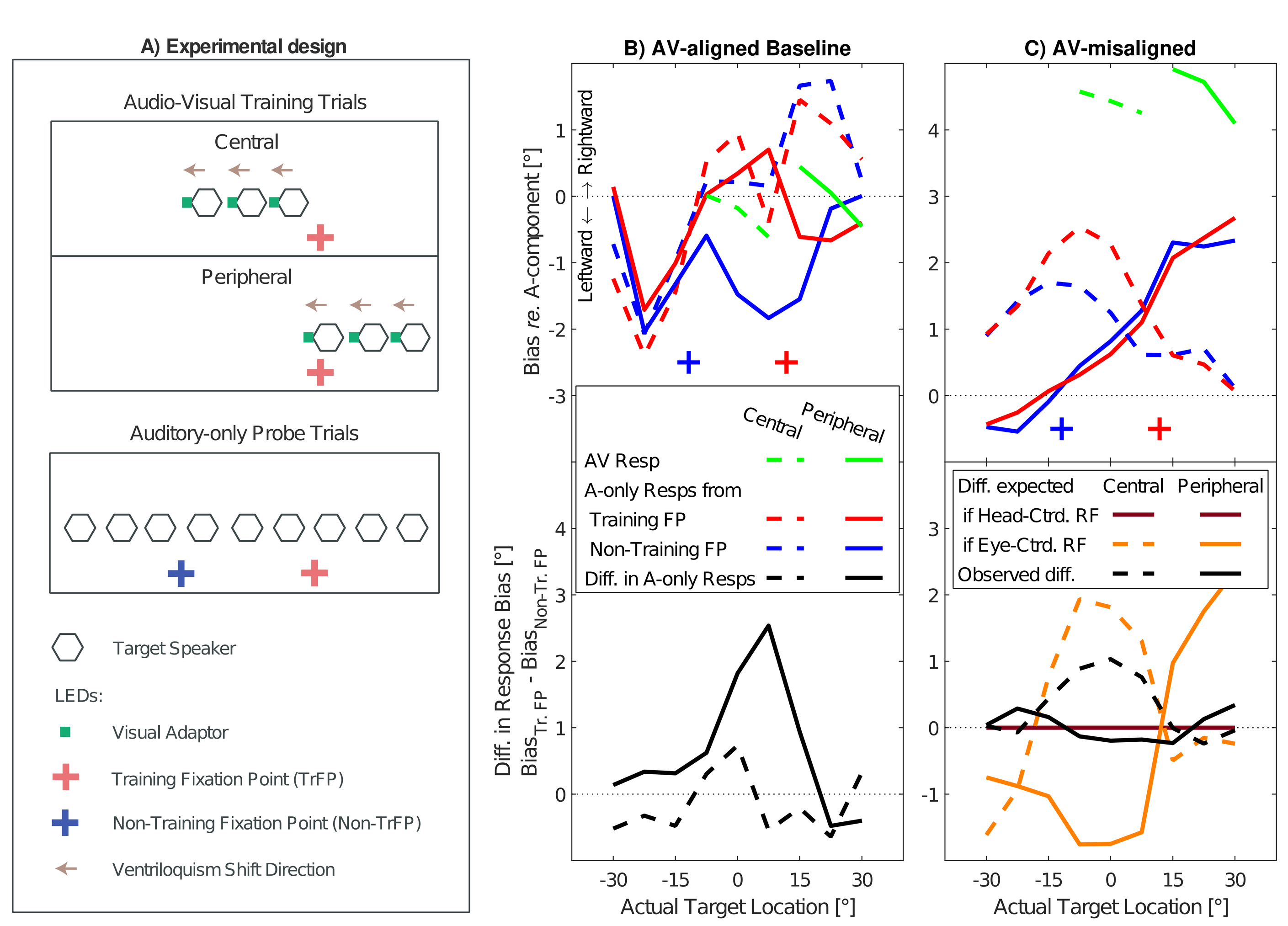


Fig. 1. A) Experimental stimuli and setup from Kopčo et. al. (2009, 2019). B) Localization bias for no-shift AV-aligned baseline condition. C) Experimental results for conditions with visual components shifted re. auditory components.

Setup and stimuli (Fig. 1A):
A stimuli: 300ms broadband noise,
V stimuli: LEDs synchronized with sound.
AV stimulus **disparity** (fixed within session):
no shift (0°); **pos.** shift (V offset +5° re A); **neg.** shift (V -5° re 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.

No-Shift Results (Fig. 1B):
Central training: responses independent of FP.
Peri. training: responses depend on FP - **unexpected form of plasticity**

Positive & Negative Shift Results (Fig. 1C):
Central training: **mix** of head- and eye-centered RFs,
Peri. training: almost **purely head-centered** RF,
→ **inconsistent results** for different training regions.

Modeling questions:
What mechanisms can explain the mixed RF of central data and the inconsistent results across central vs. peri training data?

4. MODEL WITHOUT SACCADE ADAPTATION (Lokša & Kopčo, 2023)

The **dHEC model** (Fig. 3) predicts bias in resp to an A-only target (from a fixed FP and for a given set of AV responses) as a weighted sum of:
1) Saccade-related EC bias (B in Fig 2) independent of the visual signals, caused, e.g., by hypometry of saccades and by *a priori* response bias toward the periphery,
2) Bias caused by adaptation of auditory space representation to visual signals (C in Fig. 2), defined as proportional shift towards the AV-responses, dependent on distance of the A-only target from training region. This bias is independent of properties of auditory saccades.

Four versions of the model examined, differing in optional components (gray in Fig. 2):
- **HC:** vent. signals converted to HC reference frame for adaptation (no optional comp.),
- **HEC:** vent. signals in both HC and EC RFs adapt auditory representation ("EC" arrow).
- **dHC:** similar to HC but with FP-dependent attenuation ("FP-depend. atten." comp).
- **dHEC:** similar to HEC but with FP-dependent attenuation ("EC" and "FP-dep atten" comp).

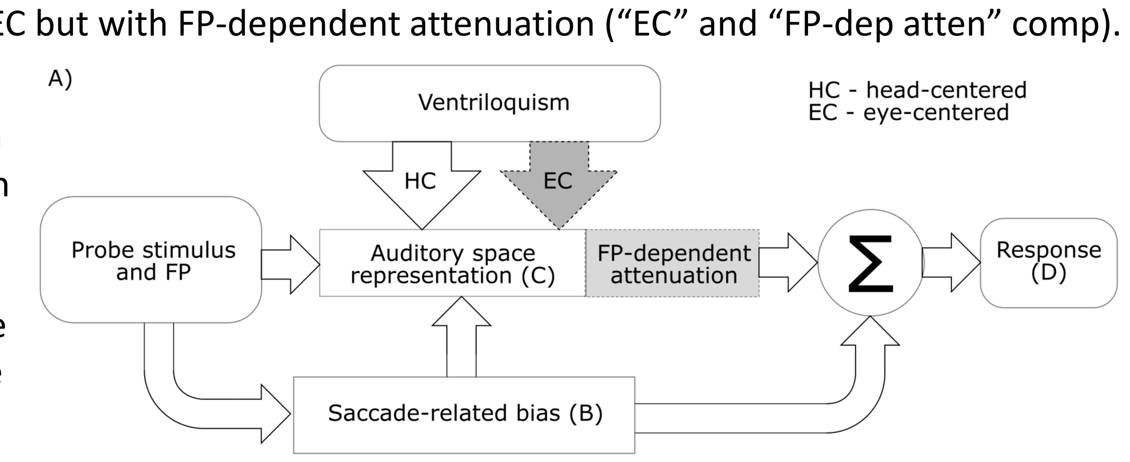
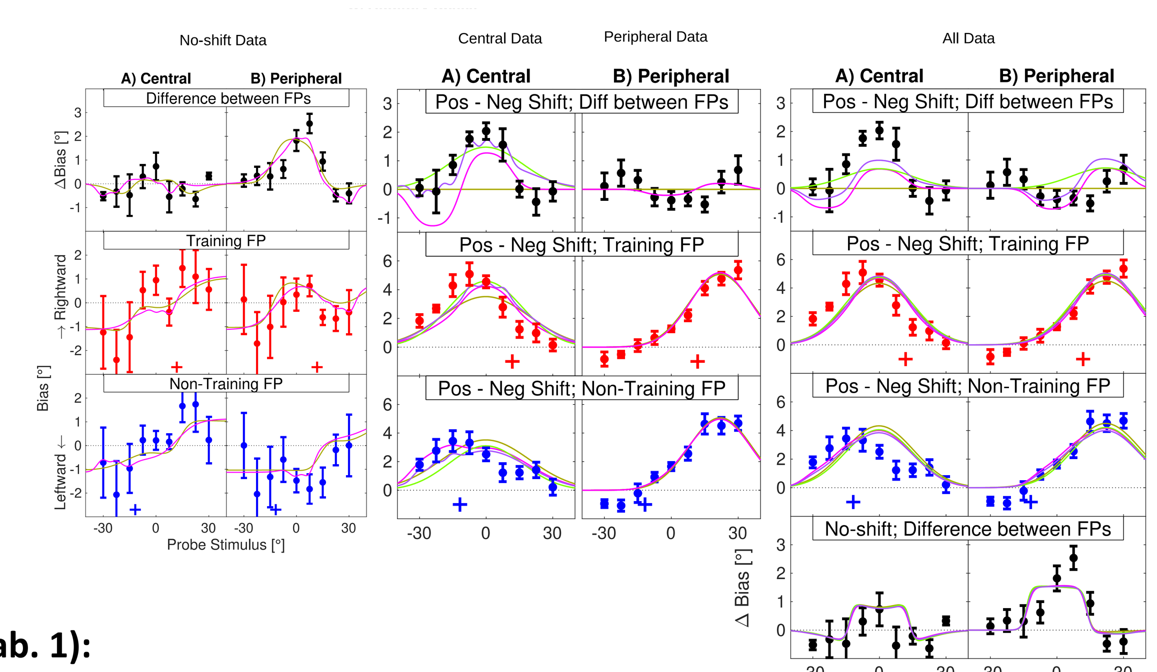


Fig. 2: Block diagram of the **dHEC** model in which optional model components that differentiate the 4 model versions are shown in gray.



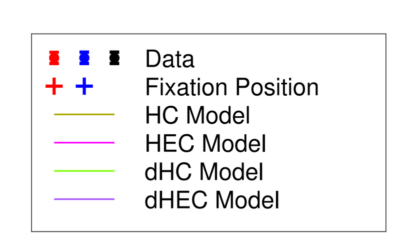


Fig. 3: dHEC model evaluations on 4 data sets. Model predictions (lines) and experimental data (symbols).

Results (also see Tab. 1):
For **no-shift data** the HC model is sufficient.
For **central data**, FP-dependent attenuation (dHC) is more likely the mechanism causing mixed RF than eye-referenced ventriloquism (HEC).
For **peripheral data**, HC model is sufficient.
For **all data** combined, FP attenuation is (dHC) helpful again, however performance is poor.

This model is able to fit no-shift data, central data, and the peripheral data separately, but not with a single set of model parameters (i.e., all data). For central and all data, it suggests that ventriloquism signals in eye-centered RF are not the likely cause of mixed RF. For no-shift data, saccade-related bias is a sufficient mechanism.

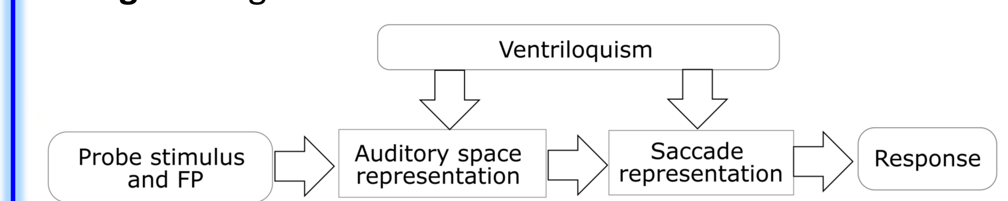
Simulation	Model	Performance		
		AICc	ΔAIC	MSE
No Shift	HC	130.9	2.4	1.59
	HEC	128.5	-	1.26
	dHC	133.8	5.3	1.59
	dHEC	131.9	3.3	1.26
Central	HC	176.2	15.6	5.48
	HEC	170.2	9.6	3.86
	dHC	160.6	-	3.22
	dHEC	162.0	1.4	2.74
Peripheral	HC	136.3	-	1.73
	HEC	141.9	5.6	1.68
	dHC	139.1	2.8	1.73
	dHEC	144.2	7.9	1.63
All Data	HC	444.7	10.5	3.25
	HEC	436.9	2.7	2.89
	dHC	436.4	2.2	2.95
	dHEC	434.2	-	2.76

Table 1: Fitted model parameters and model performance for each simulation. AICc and MSE were calculated on the data used in a given simulation. The underlined model names indicate the model version with substantial evidence of better fit to the data (i.e., round(ΔAIC) < 2).

5. MODEL WITH SACCADE ADAPTATION

The **SA-HC model** (Fig 4) focuses on explaining both central and peripheral data (Fig. 1C) using one mechanism. It assumes that auditory space is adapted by visual signals only in HC RF (like in HC version of dHEC model), while the saccades, used for responding, are also adapted – in EC RF. Specifically, **it assumes that**
- during training, **the saccades are adapted to be hypometric or hypermetric,** depending on FP, A component, and V component locations,
- during testing, **the adapted saccades either enhance or reduce the bias due to auditory space representation,** depending on the A target location vs FP.

Fig. 4: Diagram of the current model.



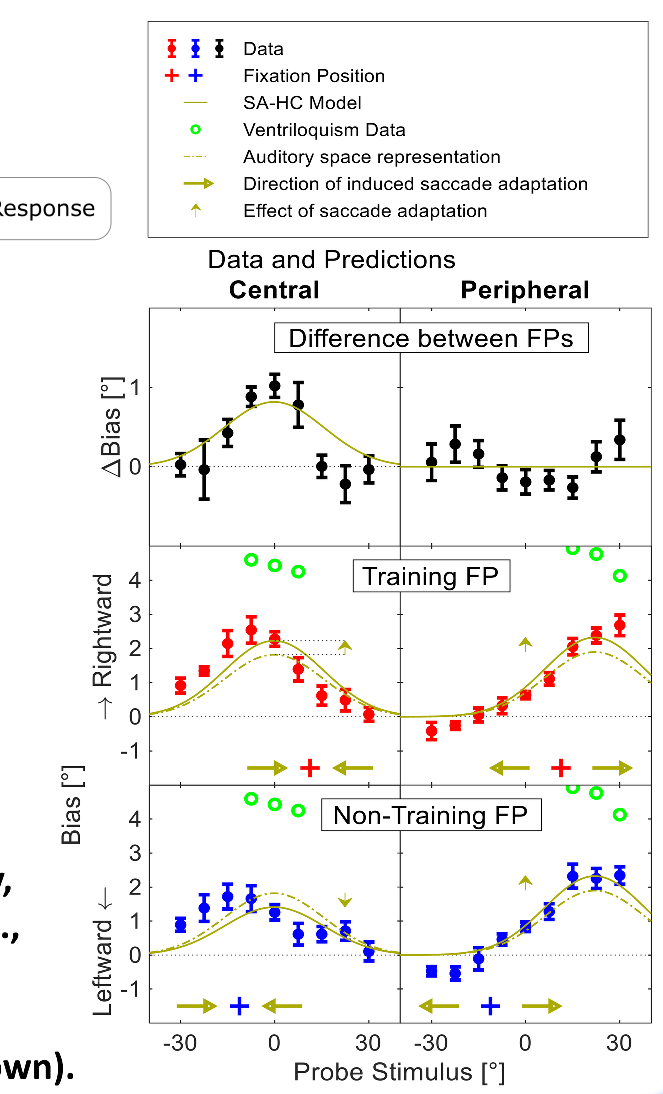


Fig. 5: SA-HC model evaluation on central & peripheral data. Model predictions (lines) and experimental data (symbols). Arrows and dash-dotted lines illustrate model operation: Ventriloquism (o) determines auditory space adaptation (---) which is additionally modified by the ventriloquism-adapted hypo/hypermetric saccades (←).

Results:
SA-HC model
- fits the central and peripheral data simultaneously,
- suggests that ref. frame of VAE is purely head-cent., and the previously observed mixed RF was due to saccade adaptation,
- cannot predict the no-shift data (of Fig 1B; not shown).

6 CONCLUSIONS AND DISCUSSION

- We introduced two models 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 can predict the newly reported adaptation by AV-aligned stimuli (Kopco et al., 2019) as a combination of saccade-related biases corrected by visual adaptation.
- A model that assumes that FP-dependent attenuation of auditory representation (dHC model) predicts the mixed RF of central data than a model with eye-referenced ventriloquism signals (HEC). → The RF might not actually be mixed.
- No version of dHEC model could describe central and peripheral data simultaneously. SA-HC model with the saccade adaptation fits central and the peripheral data simultaneously -> confirmation that RF of VAE is most likely not mixed.
- Next steps:
  - **combine** the **models** to obtain one **unified model** that can predict all data.
  - experimentally **test** the model **predictions** about **saccade-related EC bias** (Fig. 2) and **saccade** representation **adaptation** (Fig. 4), as well as the prediction that the **reference frame of the VAE is purely head-centered**.

ACKNOWLEDGMENTS

Work supported by Horizon Europe HORIZON-MSCA-2022-SE-01 grant N° 101129903 and VEGA 1/0350/22

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
Lokša P, Kopčo N. (2023). Toward a Unified Theory of the Reference Frame of the Ventriloquism Aftereffect. Trends in Hearing. 2023;27