Visually-guided Auditory Adaptation and 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 have to be aligned. A previous study examining RF of VA found inconsistent results: the RF was a mixture of HC and EC for VA induced in the center of the audiovisual field, while it was predominantly HC for VA induced in the periphery [Lin et al., JASA 121, 3095, 2007]. In addition, the study found an adaptation in the auditory space representation even for congruent AV stimuli in the periphery. Here, a computational model examines the origins of these effects. The model assumes that multiple stages of processing interact: 1) the stage of auditory spatial representation (HC), 2) the stage of saccadic eye responses (EC), and 3) some stage at which the representation is mixed (HC+EC). Observed results are most consistent with a suggestion that the neural representation underlying spatial auditory plasticity incorporates both HC and EC auditory information, possibly at different processing stages.

2. INTRODUCTION

- Ventriloquism aftereffect (VAE) – short-term change in auditory spatial perception induced by visual signals.
- Unimodal spatial reference frame (RF) in audition is head-centered, in vision eye-centered (Brainard, Knudsen, 1995; Razavi et. al., 2007).
- Is **RF** of **VAE head-centered**, **eye-centered**, or **mixed**?
- Behavioral results inconsistent: **mixed** RF found in **central** visual field, **head-centered** in **peripheral** (Lin et al, 2007).
- Understanding RF of VAE can inform us about **general properties supramodal** spatial representation in the brain.

Current study

- **Model** of ref. frame of VAE using a combination of head-centered signals, eye-centered signals, and a priori biases.
- **Analyze data** of Kopco et al. (2009) and Lin et al. (2007) to determine whether VAE has equal strength for hypometric and hypermetric saccades when VAE is assessed by eye saccades to perceived auditory target location.

3. EXPERIMENTAL DATA OF LIN ET AL. (2007)

Methods

- VAE induced with eye-gaze fixed at one fixation point (FP), called training FP (TrFP), AV stimulus shift direction constant within session (Fig. 1A).
- VAE measured from two different FPs: TrFP and non-training FP (NtrFP).
- If induced response biases **shift with FP** then RF is **eye-centered**; if response bias **do not shift with** FP, then RF is **head-centered** (Fig. 1B).
- Spatial auditory studies often find differences in results between stimulation of **central** vs. **peripheral** region (e.g., Maier et. al., 2009). Here, the effect is examined for two different training regions: central (shaded area in Fig. 2A and 2B), peripheral (three right-most speakers in lower panel of Fig. 2A)

Setup and stimuli:

- A-stimuli: 300ms broadband nosie, V-stimuli: LEDs synchronized with sound.
- AV stimulus disparity: depends on session (no shift: 0°; positive shift: V shifted 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 Fixation Point (TrFP).
- A-only stimuli presented with eyes fixated on TrFP or NonTrFP.

Positive & Negative Shift Results (Fig. 2):

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

No-Shift Results (Fig. 3):

- 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),
- <u>Unexpected form of plasticity</u> observed for central locations with peripheral training.

Modeling questions:

- What is the **origin** of the **no-shift plasticity** observed with peripheral training?
- Can this plasticity be the **reason for inconsistency in RF of VAE observed with central vs. peripheral training?**

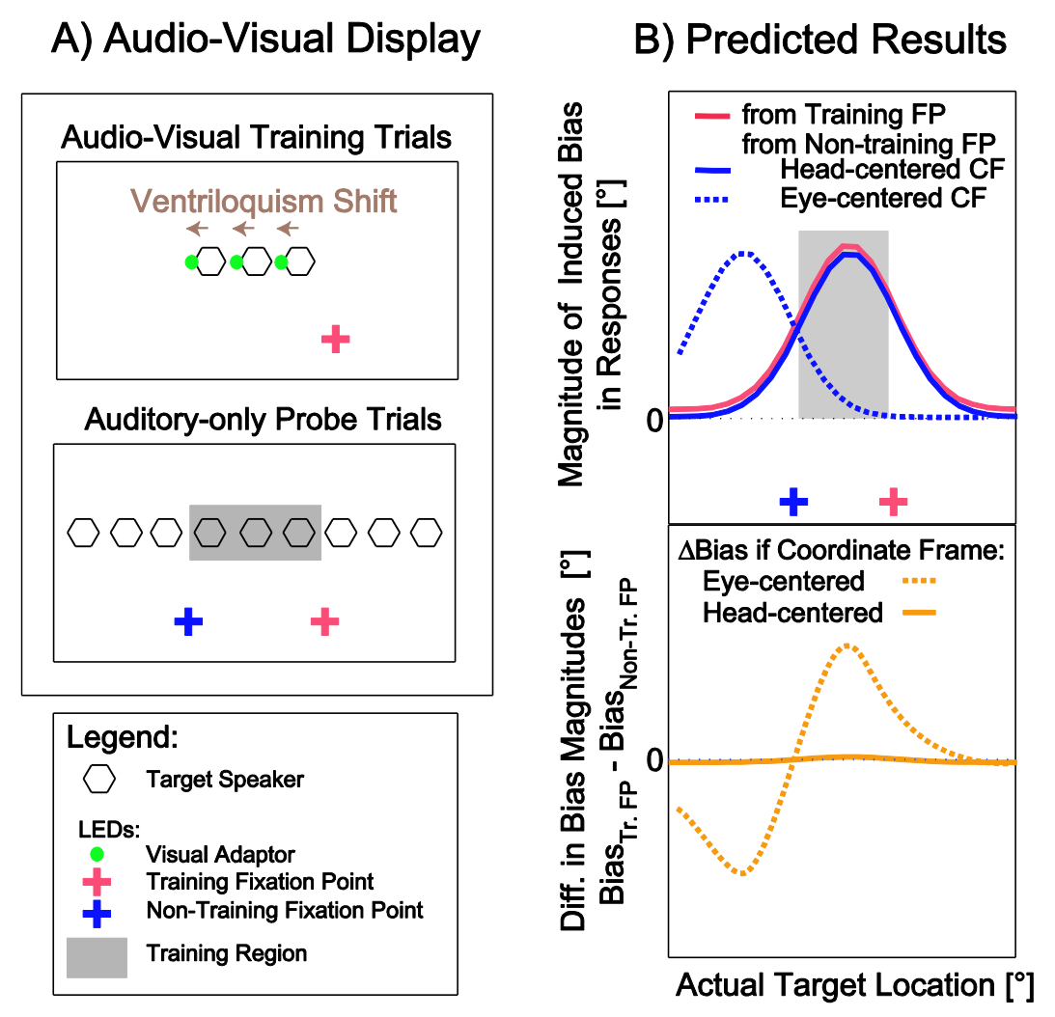


Fig. 1. Experimental stimuli and setup. Rem: setup for central training region and negative shift is illustrated here as instance. (Kopčo et. al., 2009)

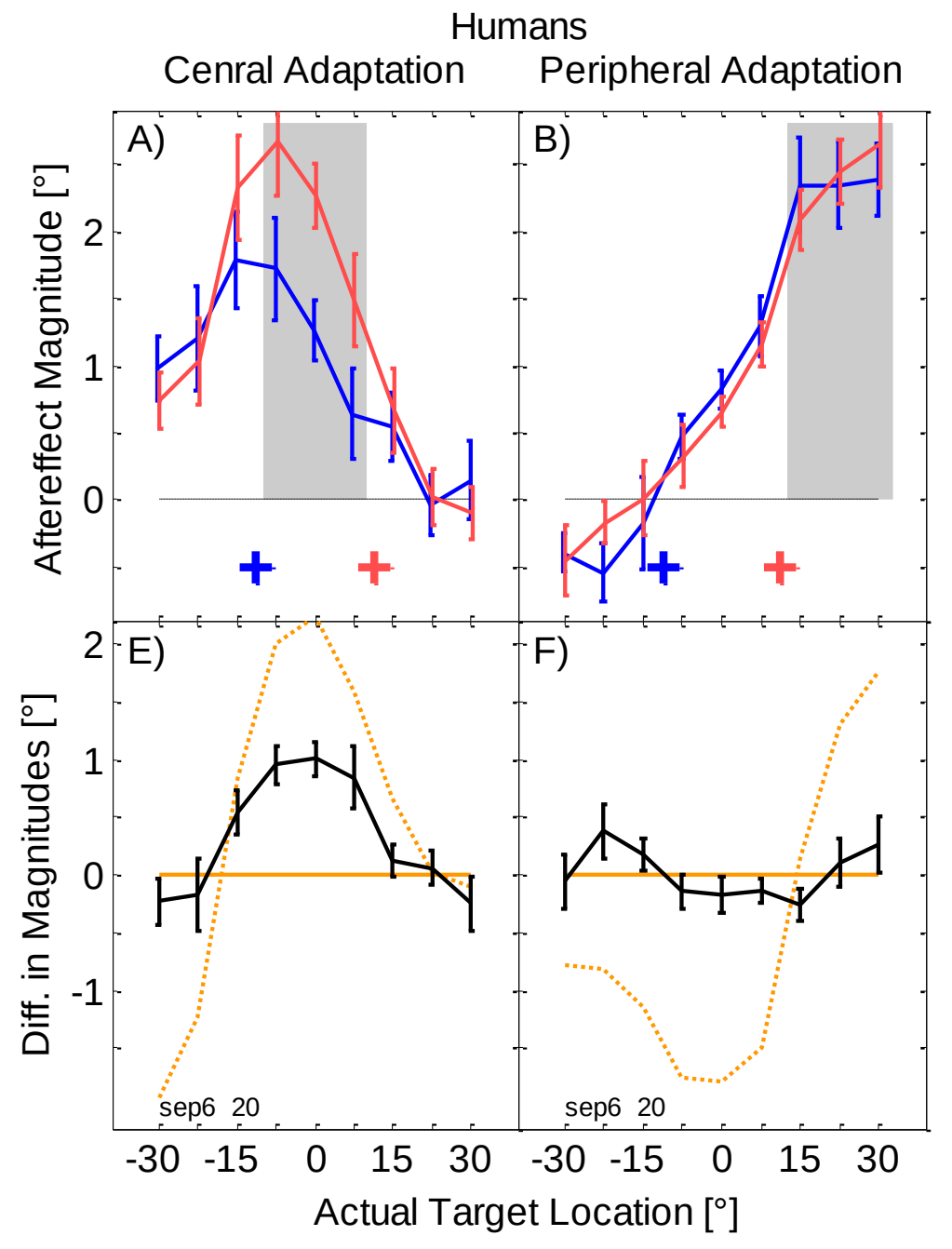


Fig. 2. Experimental results. Data for positive shift are averaged with inverted data for negative shift. Colors here are analogical to Fig. 1. Black line shows real difference in magnitude according to red and blue line in here. Yellow lines show predictions of head-centered and eye-centered model from Fig. 1. (from Kopčo et. al., 2009).

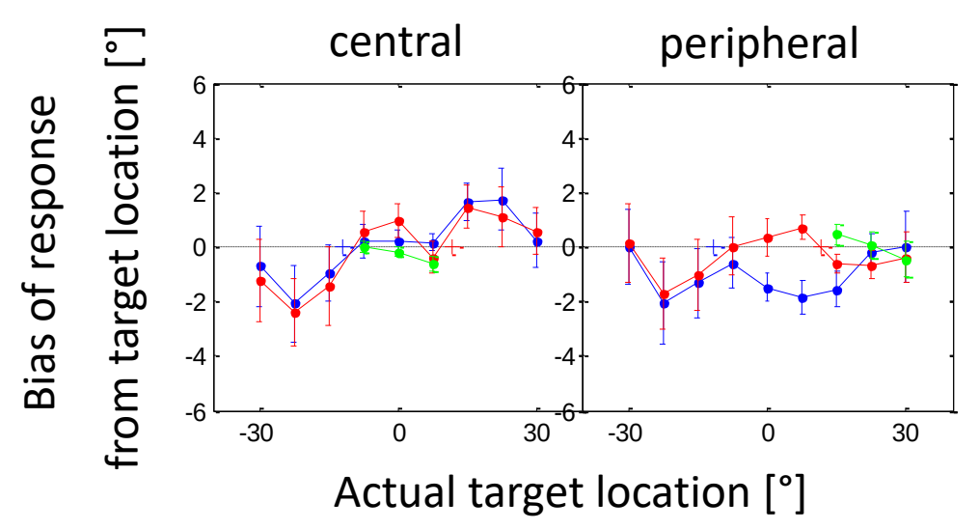


Fig. 3. Localization bias for no-shift condition according to experimental data. Colors are analogical to Fig. 1. Green color is the response to AV trials.

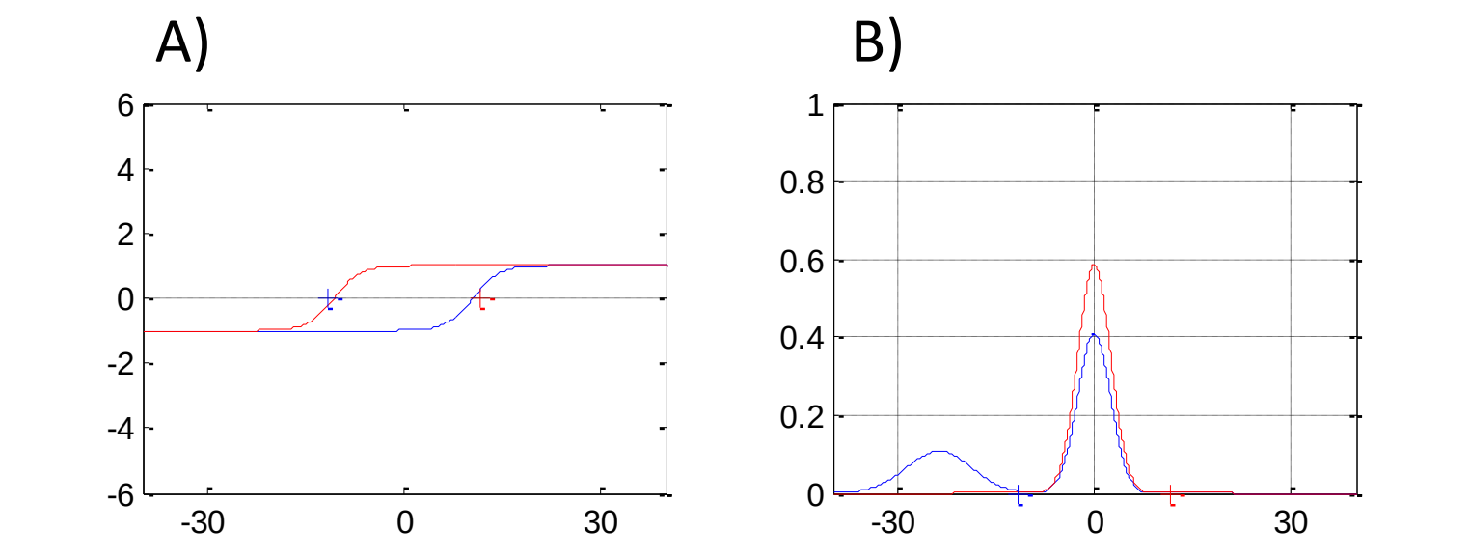
4. MODEL DESCRIPTION

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:

- A priori bias independent of the visual signals, caused, e.g., by hypometry of saccades (CITATION eg Yao and Peck), inherent bias toward the periphery (Razavi et al., 2007), and other factors, characterized by a sigmoid with 3 parameters (Fig 4A),
- Bias caused by visual signals, defined as attraction towards the AV-responses, dependent on distance of the A-only target from each AV-response. The distance is defined using a Gaussian (Fig. 4B) and aligned with the A target either using head-centered or eye-centered coordinates.

The model does not consider hypo/hypermetry in adaptation (see next section).

A priori bias function has arbitrary components that are dependent on eye position.



$$h = 1, s = 0.5, c = 0.9, w = 0.6, w_E = 0.3, \sigma_H = 2.5, \sigma_E = 5$$

Fig. 4: Components of the model: A) A priori bias function for the TrFP and NonTrFP is defined by h (height), s (slope), and c (center-offset) of a sigmoid. B) Neighborhood in which a given AV stimulus influences the responses to A-only targets is defined by a Gaussian centered on AV signal with width of σ_H or σ_E. The strength of the AV target's influence on the A-target is assessed in head-centered (σ_H) and eye-centered coordinates (σ_E), w_E defines relative weight of the two coordinates.

5. IS VENTRILOQUISM AFTEREFFECT SIZE DEPENDENT ON SACCADE DIRECTION?

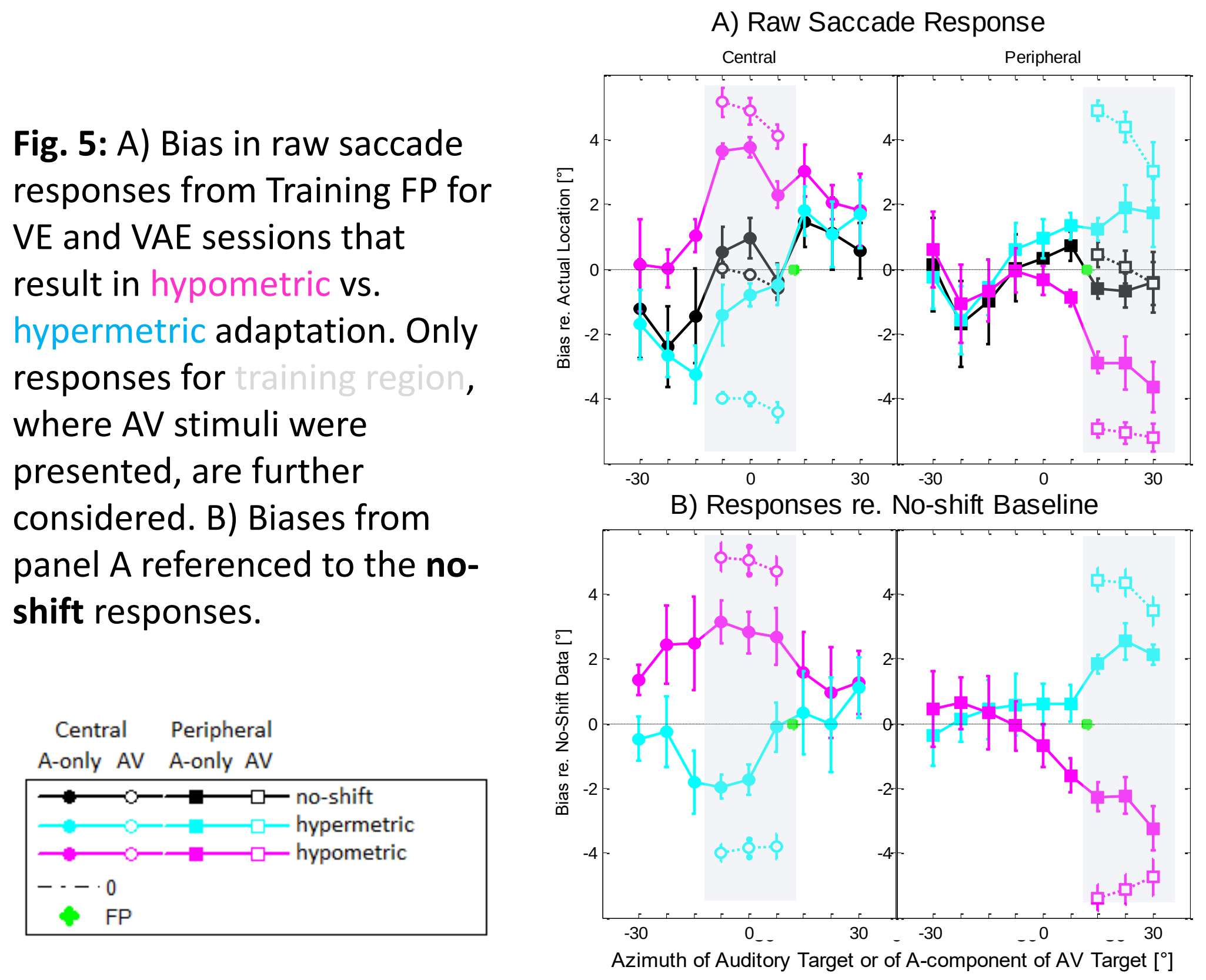
Lin et al. (2007) studies were performed such that eye saccades were used both to induce ventriloquism and to evaluate VE and VAE. Saccades use eye-centered RF. Therefore, if asymmetrical VAE is induced for conditions resulting in hypometric vs. hypermetric saccades, that might contribute to the observed eye-centered component of the RF of VAE.

Fig. 5A shows that, when raw biases are considered, the hypometric (eg. right-ward ventriloquism shift for a saccade that goes to the left) VAE much **stronger** than the hypermetric VAE (filled symbols in training region). **Fig. 5B** shows that a part of that asymmetry is due to shifts in the **no-shift** baseline. However, even if that is accounted for, the hypometric VAE is still stronger than hypermetric VAE.

Fig. 6A shows that the asymmetry is also present in the ventriloquism effect, such that hypometric VE is approx. 100%, while hypermetric VE is around 80%. When the VAE data are expressed, as a proportion of VE (**Fig. 6B**) VAE is approximately 50% in all conditions (except for 1 data point).

For modeling purposes, VEA can be considered as independent of saccade shift direction, as long as VE saccades are used as reference.

Fig. 5: 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) Biases from panel A referenced to the **no-shift** responses.



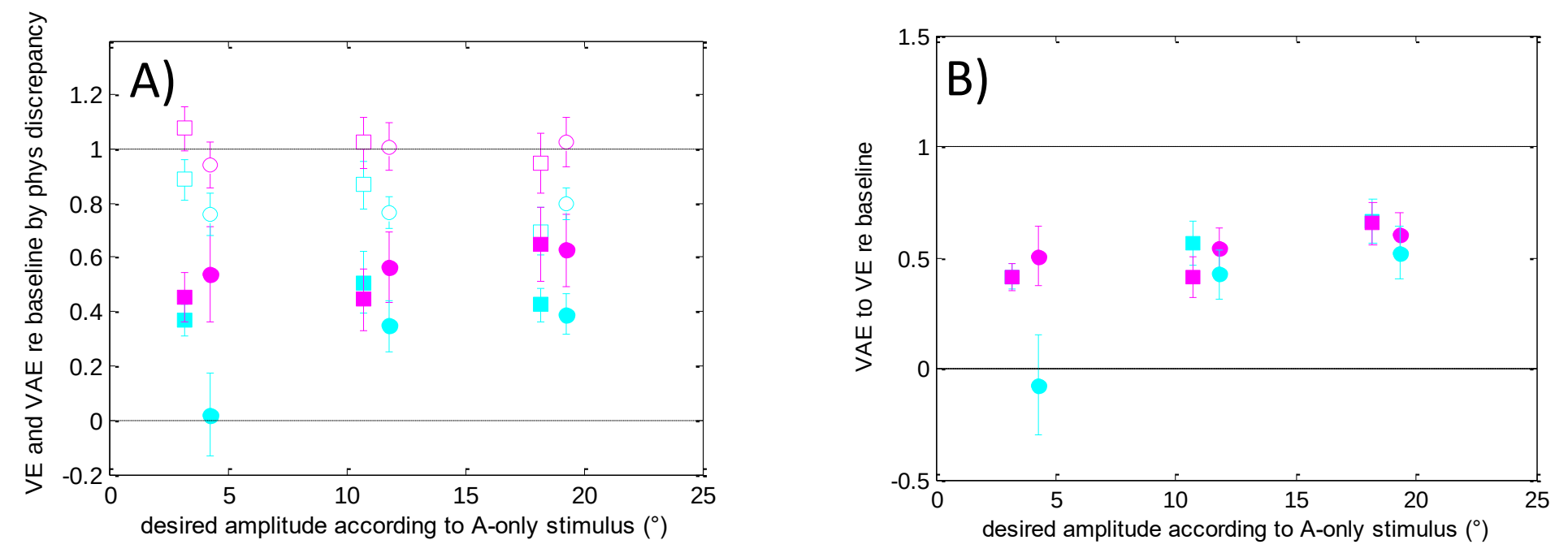


Fig. 6: Relative strength of VE and VAE hypometric and hypermetric adaptation as a function of desired amplitude (i.e., distance from FP to A-target). A) AV and A-only data from Fig. 4B scaled by the physical AV disparity. B) VAE as a proportion of VE from panel A.

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

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

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 (KUZ XVI).* 101-106

ACKNOWLEDGMENTS

Work supported by VEGA 1/1011/16 and EU H2020-MSCARISE-2015 Grant No. 69122.

6. MODELING RESULTS

No-shift data from center and periphery can be fitted very well (Fig. 7A).

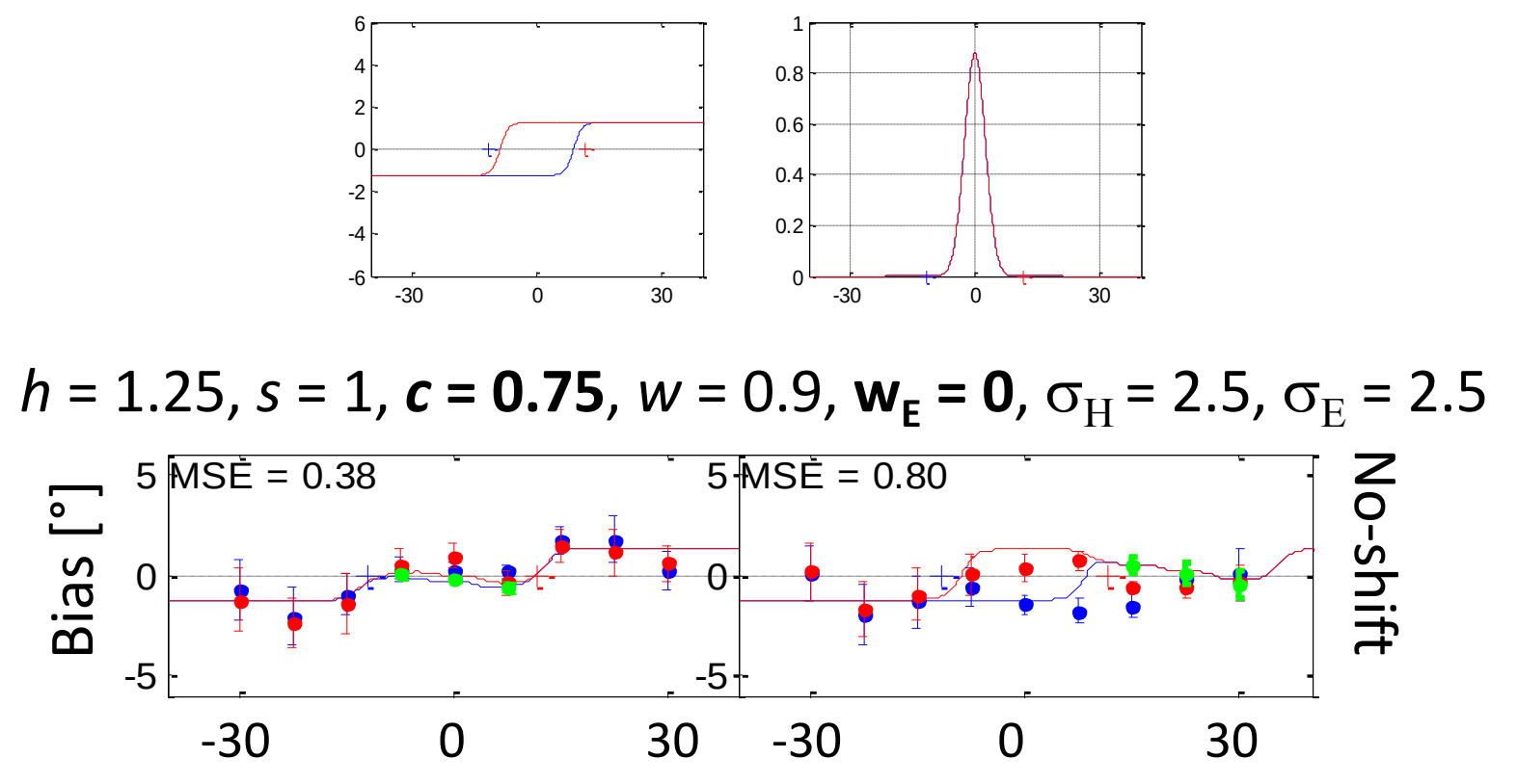
Shift data for Center and fitted well if eye-centered representation is strongly weighted (Fig. 7C).

Shift data for Periphery are fitted well if eye-centered a priori component is present (Fig. 7D).

A combined fit of Shift and No-shift data for Center and Periphery also require eye-centered and head-centered representation (Fig. 7B).

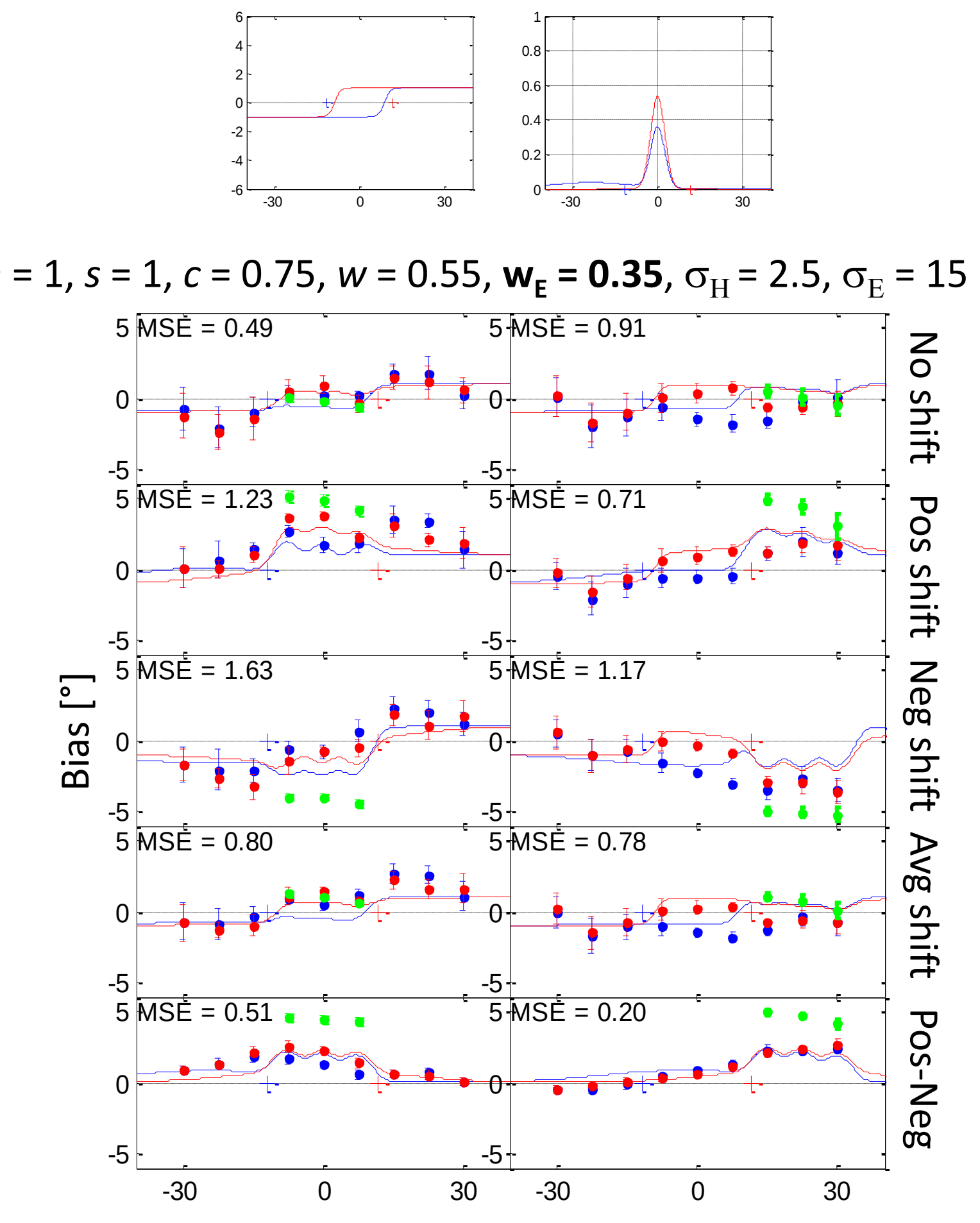
Fig. 7: Performance of the model fitted to different subsets of data. Panels show the fitted a priori bias and Gaussian neighborhood function, and predictions for central (left column) and peripheral (right column) data for different shift conditions (rows).

A) No-shift Central and Peripheral Data



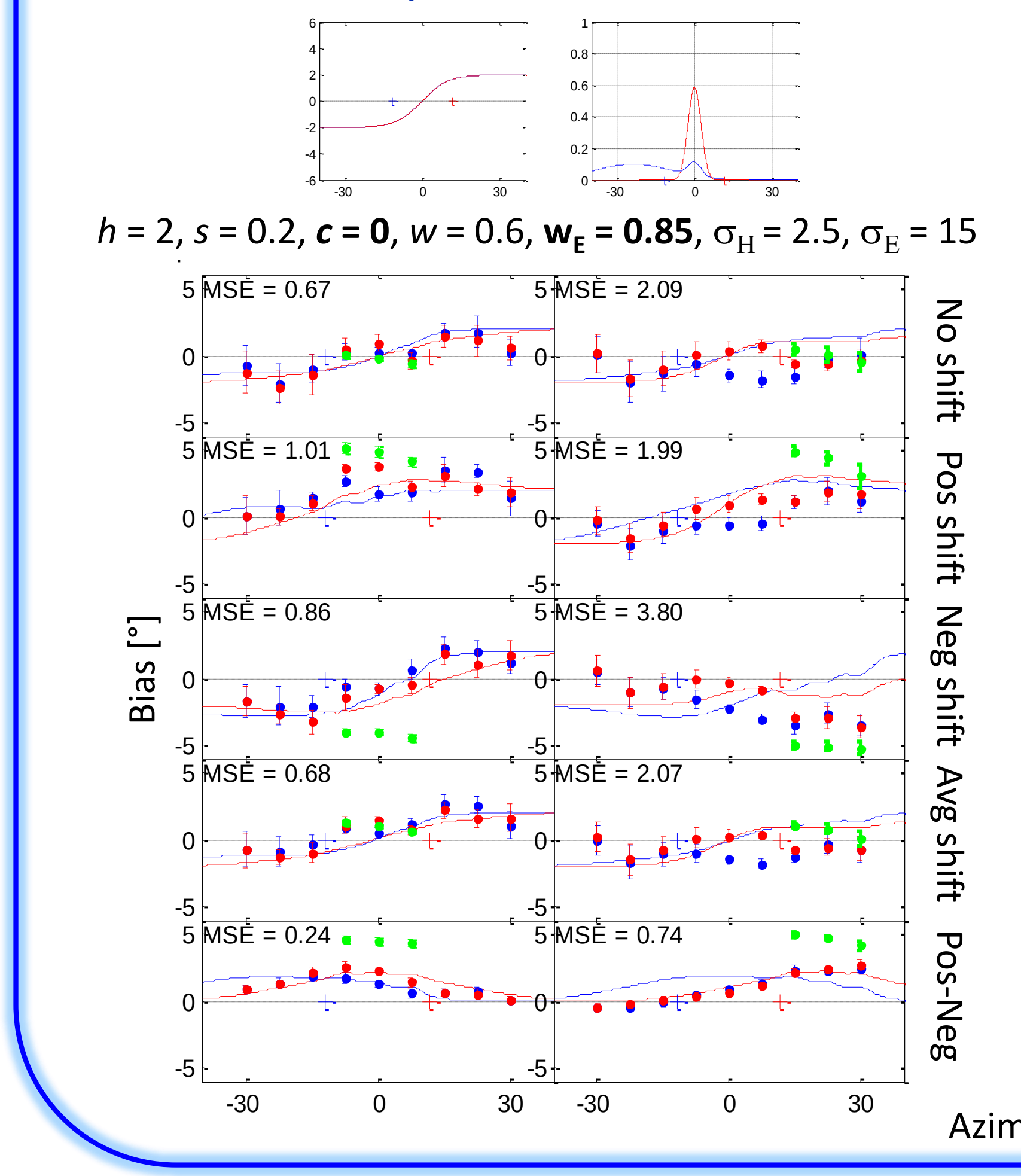
h = 1.25, *s* = 1, ***c* = 0.75**, *w* = 0.9, $\mathbf{w_E = 0}$, $\sigma_H = 2.5$, $\sigma_E = 2.5$

B) Shift and No-shift data for Center and Periphery



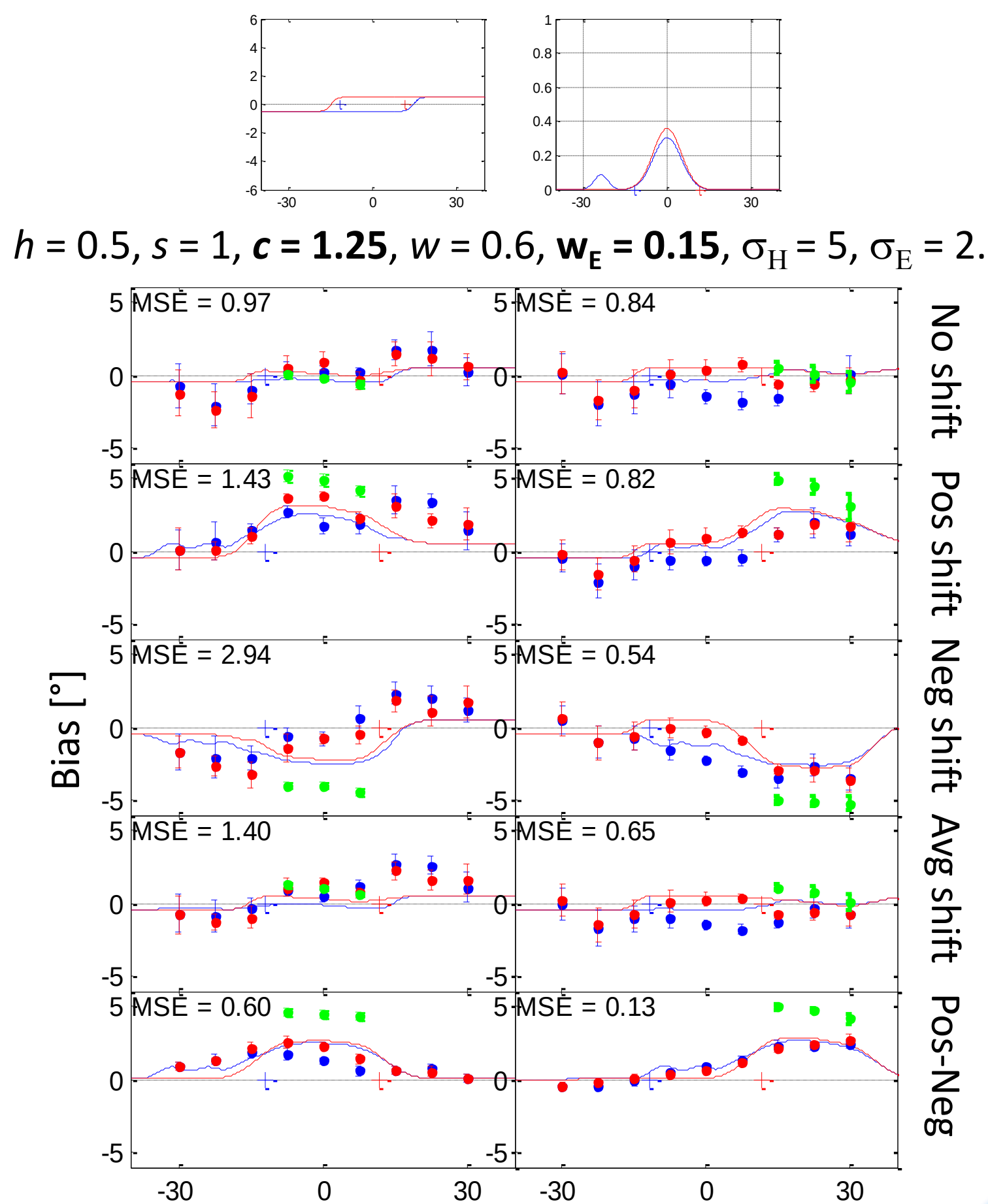
h = 1, *s* = 1, *c* = 0.75, *w* = 0.55, $\mathbf{w_E = 0.35}$, $\sigma_H = 2.5$, $\sigma_E = 15$

C) Shift Data for Center



h = 2, *s* = 0.2, ***c* = 0**, *w* = 0.6, $\mathbf{w_E = 0.85}$, $\sigma_H = 2.5$, $\sigma_E = 15$

D) Shift Data for Periphery



h = 0.5, *s* = 1, ***c* = 1.25**, *w* = 0.6, $\mathbf{w_E = 0.15}$, $\sigma_H = 5$, $\sigma_E = 2.5$

7. CONCLUSIONS AND DISCUSSION

- Ventriloquism is stronger if resulting in hypometric saccades (vs. hypermetric) when saccades to sounds are used as response measure, both for Ventriloquism Effect and Aftereffect. Considering VAE as a proportion of VE eliminates this asymmetry.
- Model that considers eye-centered and head-centered representation is required to describe the data, suggesting that the **reference frame of Ventriloquism aftereffect is mixed** both in center and in periphery, at least when eye-saccades are used as response measure.
- **Eye-centered** contribution to the mixed representation is always broader in the modeling results -> it does not seem to be a simple shifted copy of the **head-centered** signal.
- Eye-centered dependence in the peripheral no-shift data can be explained by assuming a form of eye-centered a priori bias that is eliminated by correct AV signals. However, it is not clear whether this bias is a result of saccade adaptation or auditory representation adaptation.
- Current overall fit of model produced smaller errors for peripheral data, resulting in underestimation of the eye-dependence of the central data. Equal weighting of the two data sets might produce an even stronger eye-dependence.
- Audio-visual integration requires multiple representations and transformations for good representation of multimodal environment.