A model of the reference frame of the ventriloquism aftereffect based on head-centered, eye-centered and distance-dependent signals

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 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. Here, a computational model introduced in (Kopco & Loksa, 2021) is extended to examine these inconsistencies, assuming that there is a fixed relationship between the VAE and the ventriloquism effect.
Methods: The model has two components: a saccade-related component characterizing the adaptation in auditory-saccade responses and auditory space representation adapted by ventriloquism signals in a combination of head-centered and eye-centered frames, in which the strength of adaptation can be eye-gaze-direction dependent. There were 4 different model versions implemented, differing in 2 aspects. The first aspect is whether the ventriloquism aftereffect was a mix of head- and eye-centered (HEC), or purely head-centered (HC). The second aspect is whether the gaze-direction-dependent modulation was considered (dHEC or dHC) or not (HEC or HC). The model versions were compared using AICc criterion in 4 different simulations using different data sets: no-shift, all data, central and peripheral.
Results: Experimental data analysis confirmed that the VAE measured using saccades can be predicted based on observed ventriloquism effect. Overall, the model performed best when eye-centered signals were combined with head-centered signals with a gaze-direction-dependent modulation (dHEC) for all data simulation. However, for no-shift simulation where just data affected by aligned audiovisual pairs were selected, the HEC model provided the best fit to the data.
Conclusion: There are likely to be two mechanisms by which visual signals are realigned with auditory signals. These mechanisms are combined to visually calibrate the auditory spatial representation in a mixed reference frame.

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 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. 3) whether the signals keep their overall strength when eyes move to a new FP location from training FP location, or whether their strength is attenuated.
- 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)

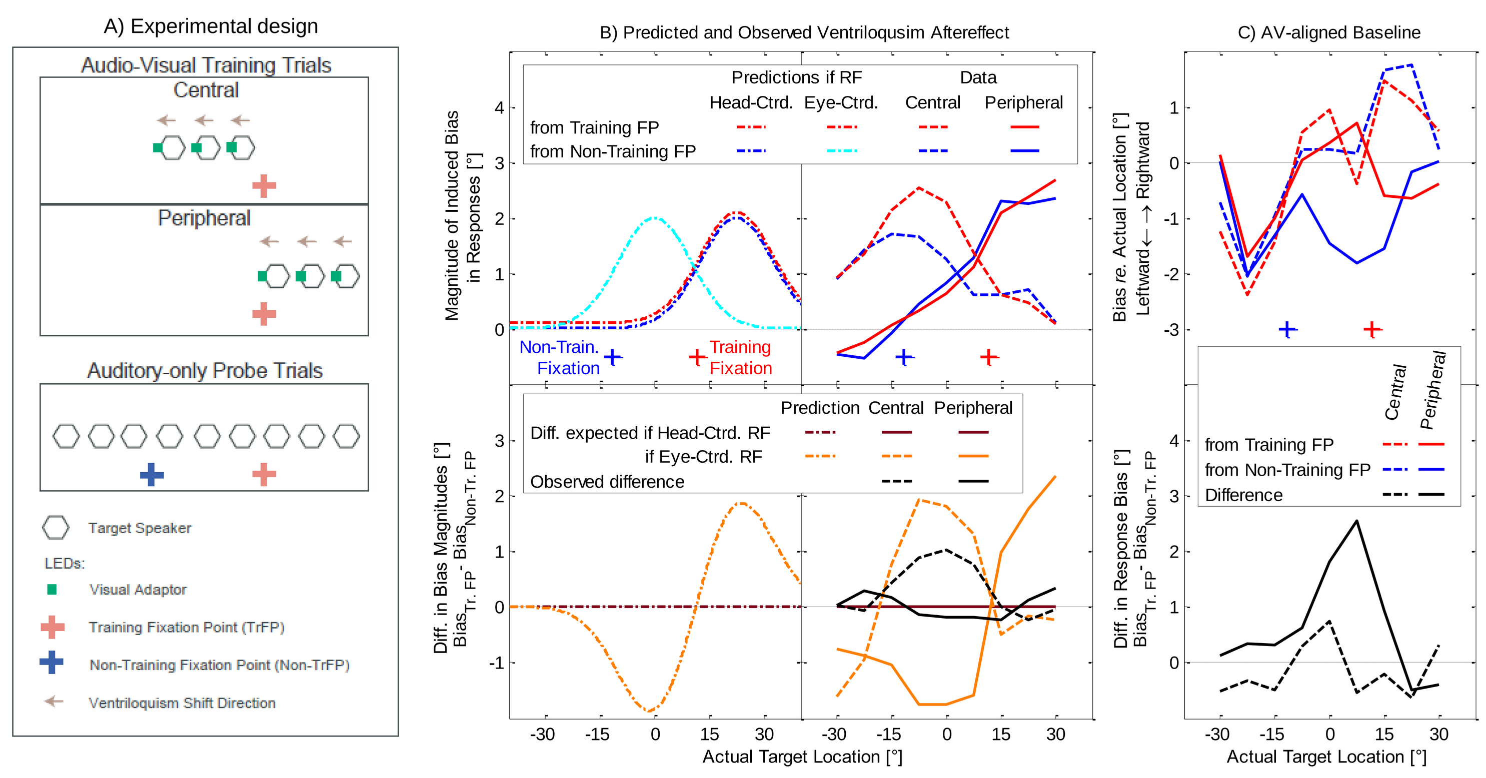


Fig. 1. A) Experimental stimuli and setup from Kopčo et. al. (2009, 2019). B) Experimental results for conditions with visual components shifted re. auditory components. C) Localization bias for no-shift AV-aligned baseline condition.

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 experiments, 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 noise, 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 signals are also considered, and thus that the RF is indeed mixed?**
- **What is the form of the EC signals? Two forms explored: EC component or FP-dependent attenuation.**

4. 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.

Four 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.
- **dHC, dHEC**: same as HC and HEC respectively except that FP-dependent attenuation is implemented here

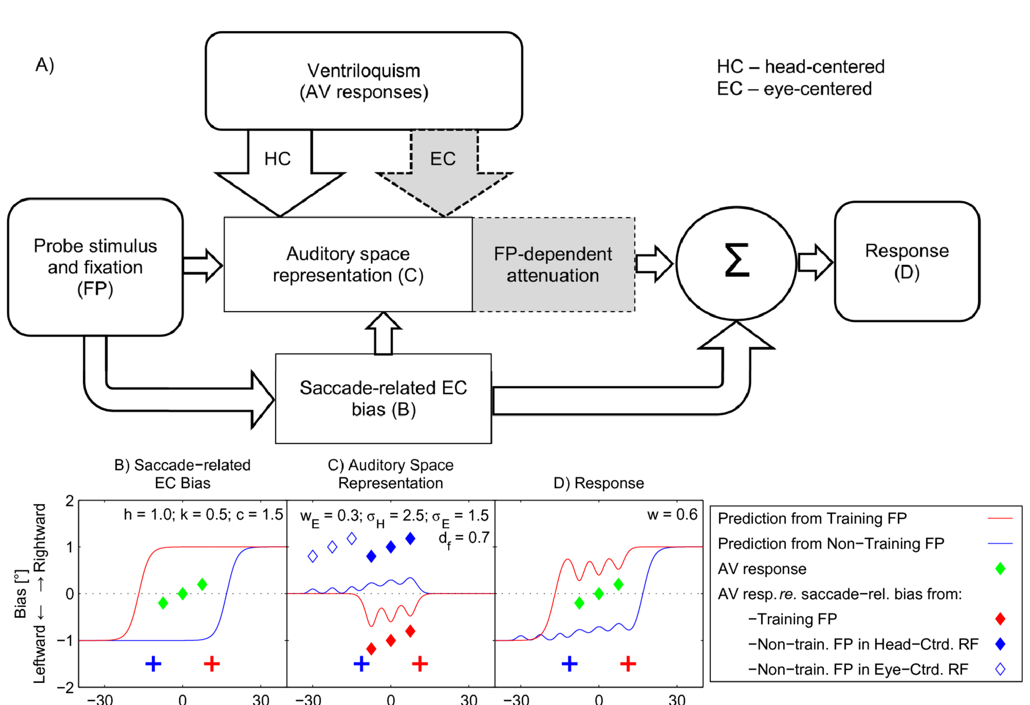


Fig. 2: A) **Model Diagram.** Response to auditory stimuli is predicted, depending on the FP location, as a linear combination of **saccade-related EC bias** (B) and **adaptation in auditory space representation** induced by the ventriloquism and proportional to the VE in the AV responses (C). The ventriloquism signals are either exclusively in HC RF (model versions HC and dHC) or in a mixture of HC and EC RFs (modesl HEC and dHEC), 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. Parameter d_f defines fraction of auditory space representation component for the predictions of the responses to the non-training fixation. **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).

5. MODELING RESULTS

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

No-shift data (Fig. 3A): The d models have no difference in prediction in comparison with the non-d models, if we compare HC with dHC and HEC with dHEC. However, because the non-d models are simpler than the d models, their AICc is better due to the model complexity penalization implemented in the AICc evaluation. HEC and HC models are comparable in terms of AICc for both non-d and d comparisons. Also, the ratio of between the strengths of EC vs. HC signals is 0.36 : 0.64 (because the $w_E = 0.36$).

→ **Adaptation in saccadic responses can explain FP-dependence of no shift data.**

All data (shift and no-shift data for center and periphery Fig. 3B):
the two "d" models are comparable to each other, slightly better than HEC. All three models are considerably better than HC model.
→ **FP-attenuation or EC component are needed to account for the data. Both of these components are EC-referenced, so EC-referenced signals are used in VAE.**

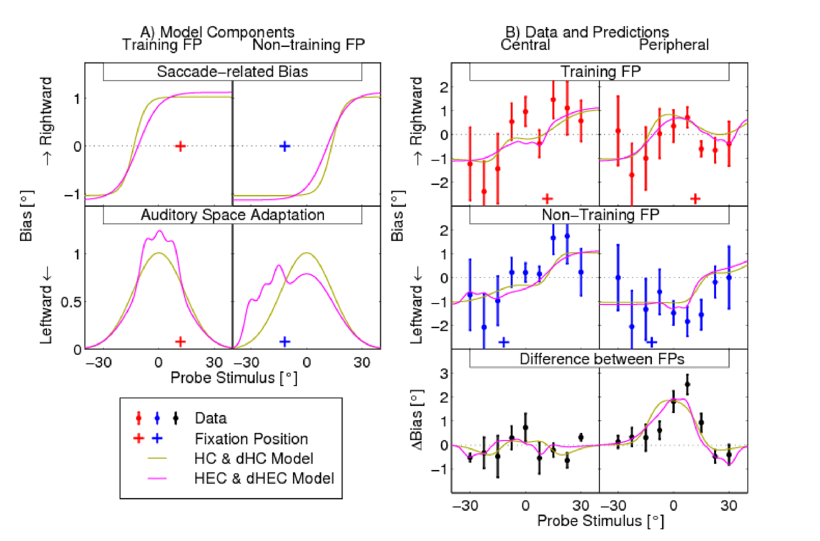
Central-training aftereffect data (Fig. 3C): Central-training data are much better described if FP-attenuation is included (dHC and dHEC models are comparable). If it is not, then the EC component is required (HEC model better than HC).
→ **FP-attenuation is needed to model the data.**

Peripheral-training data (not shown): require neither FP-attenuation nor EC component (HC model considerably better than any other).

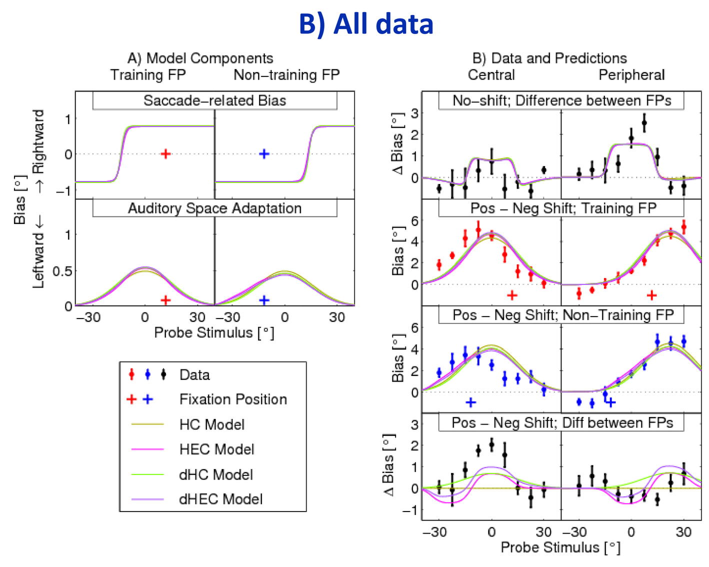
Parameters:
- w_E always less than 0.5 → EC RF contribution weaker than HC.
- σ_E always narrower than σ_H. → EC RF signals more specific than HC.
- *w* between ~0.5 and 1 → both model components important.
- d_f close to 1 → almost no FP-attenuation.

A) No-shift Data

Figure 3: 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. Error bars represent across-subject standard errors of mean (N = 7).



B) All data



C) Aftereffect for Central Data

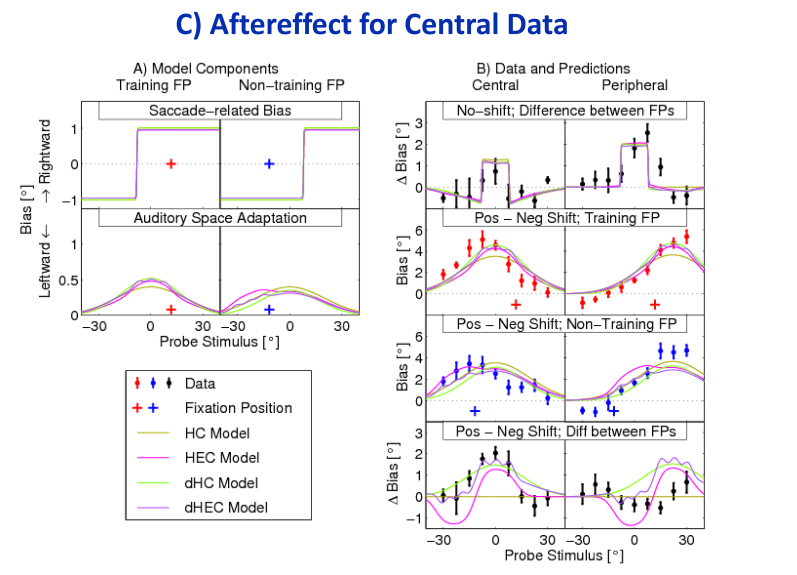


Table 1: Resulting parameter values and AICc related evaluation.

Simulation	Model	Fitted parameter values								Performance		
		H	k	c	w	w_E	σ_H	σ_E	d_f	AICc	ΔAIC	MSE
No Shift	HC	1.03	0.31	1.14	1.01	-	12.06	-	-	130.9	2.4	1.59
	HEC	1.13	0.17	0.95	1.24	0.36	12.84	2.98	-	128.5	-	1.26
	dHC	1.03	0.31	1.14	1.01	-	12.06	-	1.00	133.8	5.3	1.59
	dHEC	1.13	0.17	0.95	1.24	0.36	12.84	2.98	1.00	131.9	3.3	1.26
All Data	HC	0.79	0.82	1.15	0.49	-	14.21	-	-	444.7	10.5	3.25
	HEC	0.77	0.76	1.13	0.53	0.15	13.35	4.83	-	436.9	2.7	2.89
	dHC	0.79	0.91	1.17	0.54	-	13.74	-	0.85	436.4	2.2	2.95
	dHEC	0.77	0.82	1.15	0.55	0.11	13.64	4.36	0.90	434.2	-	2.76
Central	HC	1.01	5.64	0.67	0.40	-	18.79	-	-	176.2	15.6	5.48
	HEC	0.96	5.60	0.67	0.48	0.30	18.14	5.01	-	170.2	9.6	3.86
	dHC	1.01	6.84	0.67	0.52	-	15.20	-	0.68	160.6	-	3.22
	dHEC	0.96	13.46	0.67	0.51	0.17	17.99	2.65	0.74	162.0	1.4	2.74
Peripheral	HC	0.83	3.40	1.33	0.55	-	12.43	-	-	136.3	-	1.73
	HEC	0.82	5.33	1.33	0.56	0.04	12.12	4.91	-	141.9	5.6	1.68
	dHC	0.83	3.27	1.33	0.55	-	12.43	-	1.00	139.1	2.8	1.73
	dHEC	0.82	2.85	1.33	0.56	0.04	12.12	4.91	1.00	144.2	7.9	1.68

6. 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. It can also predict well the Peripheral data.
- The FP-dependent attenuation (dHC and dHEC models) provided best fit to Central data → **FP-attenuation is needed to model the data.**
- 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**. However, the overall fit still does not describe the differences between the data well.
- Ventriloquism adaptation considered here is local. → It is inconsistent with the models based on the opponent processing channels (Grothe et al, 2010).
- Future steps: (1) Add a component related to saccade hypo-/hypermetry to the modelling. (2) Experimentally test the prediction that saccade-related EC bias occurs and influences RF of VAE measurements.

ACKNOWLEDGMENTS

This work was supported by Science Grant Agency of the Slovak Republic VEGA 1/0355/20 and by Danube Region Strategy & The Slovak Research and Development Agency DS-FR-19-0025.

REFERENCES

Alais, D. and Burr, D., (2004). The ventriloquist effect results from near-optimal bimodal integration. Curr Biol. 14:257-262
Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of sound localization in mammals. Physiol Rev 90, 983–1012 (2010).
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
Loksa, P., Kopčo, N.(2021) A model of the reference frame of visual calibration of auditory space, VIRTUAL 44th Annual MidWinter Meeting of the Association for Research in Otolaryngology
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