A model of the reference frame of visual calibration of auditory space

Peter Loksa, Norbert Kopco

Institute of Computer Science, P. J. Safarik University, Kosice, 04001, Slovakia

Background:

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.

Methods:

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 1^{st} 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.

Results:

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.

Conclusions:

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.

[Work supported by VEGA-1/0355/20]