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

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 is introduced 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 mixed 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.

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.