AutoHPP: An automated system for head-turn preference procedure using a deep-learning-based infant head-turn classifier

Dongjin Lee[1], Michael C. Frank[2], Jae-Hun Jung[1,3], Eon-Suk Ko[4]

[1]Graduate School of Artificial Intelligence, POSTECH, South Korea.
[2]Department of Psychology, Stanford University, USA.
[3]Department of Mathematics, POSTECH, South Korea.
[4]Department of English Language and Literature, Chosun University, South Korea.

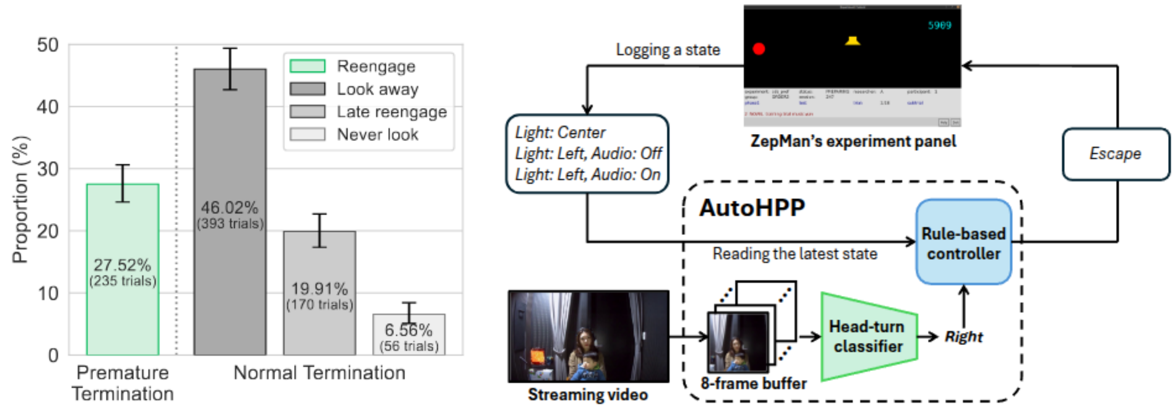
*Corresponding author(s). E-mail(s): eonsukko@chosun.ac.kr;
Contributing authors: dongjinlee@postech.ac.kr; mcfrank@stanford.edu;
jung153@postech.ac.kr;

The Head-turn Preference Procedure (HPP; Kemler Nelson et al., 1995) has been widely used to measure infants' auditory preference. In HPP, the experimenter controls stimulus presentation in *real time*, contingent on the infant's orientation toward a sound source: the experimenter initiates a stimulus and, if necessary, triggers a halt signal when the infant looks away, after which the system waits up to two seconds. If the infant reorients within this window, the signal must be canceled; otherwise, the trial ends. Delayed responses may prematurely terminate the trial despite the infant reorienting toward the stimulus, potentially disrupting the contingency between infant behavior and stimulus presentation and introducing measurement noise. Despite these concerns, prior work has rarely quantified how often such premature terminations occur. Consequently, unlike eye movements, where automated annotation is now common (Werchan et al., 2023; Erel et al., 2023), efforts to automate head-turn coding remain limited. This study addresses these gaps: we first quantify the incidence of premature terminations in an existing dataset, and we develop AutoHPP, an automated system for HPP using a deep-learning-based head-turn classification model.

We utilized an in-house Korean HPP dataset, comprising 194 video recordings. Frame-level labels (Left, Right, Center, Uncertain) were assigned a posteriori by three trained coders with high inter-rater reliability (Fleiss' $\kappa = .831$). In this dataset, we identified 1,386 trials based on an onset detection algorithm. Among these, 854 were terminated by the 2-s rule and further categorized into four groups: Reengage, Late reengage, Look away, or Never look. Reengage corresponds to premature termination and the other three to normal termination. Premature terminations accounted for 27.52% of the trials (Fig. 1a), suggesting substantial room for improvement in real-time coding accuracy. A logistic mixed effects model (Premature termination ~ Head-turn frequency + Age + (1 | Experimenter) + (1 | Infant)) showed that head-turn frequency robustly predicted premature termination ($\beta = .246$, $p < .001$). This suggests that increased behavioral variability poses a significant challenge to an experimenter's real-time judgment of infants' engagement.

AutoHPP consists of two components: (1) a CNN-LSTM model trained to classify the infant's head-turn in real time and (2) a rule-based controller that automatically runs HPP based on head-turn predictions (Fig. 1b). The head-turn classifier was evaluated on hold-out validation sets with human annotations (Cohen's $\kappa = .91$), which demonstrates that our classifier can accurately predict infants' head-turns. To evaluate AutoHPP in live testing, we compared AutoHPP data from 10 infants (15 tested, 5 excluded) against a previously collected Korean ManyBabies1 dataset obtained under human operation (The ManyBabies Consortium, 2020).

Although AutoHPP yielded longer overall listening times than human operation (OperationMode: $p < .01$), consistent with earlier trial termination by human operators, there was no evidence that operation mode altered the effect of trial type (TrialType × OperationMode: $p = .71$). These preliminary findings suggest that AutoHPP reproduces the substantive conclusions of the ManyBabies1 paradigm while reducing reliance on human operators.



(a) (b)

Figure 1. (a) Proportion of trials terminated by the 2-second rule (b) Overview of AutoHPP, an automated HPP system that converts live video into real-time control decisions through a CNN-LSTM classifier and a rule-based controller.

Reference

Erel, Y., Shannon, K. A., Chu, J., Scott, K., Kline Struhl, M., Cao, P., Tan, X., Hart, P., Raz, G., Piccolo, S., Mei, C., Potter, C., Jaffe-Dax, S., Lew-Williams, C., Tenenbaum, J., Fairchild, K., Bermano, A., & Liu, S. (2023). iCatcher+: Robust and Automated Annotation of Infants' and Young Children's Gaze Behavior From Videos Collected in Laboratory, Field, and Online Studies. Advances in Methods and Practices in Psychological Science, 6(2), 25152459221147250. https://doi.org/10.1177/25152459221147250

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. Infant Behavior and Development, 18(1), 111–116. https://doi.org/10.1016/0163-6383(95)90012-8

Werchan, D. M., Thomason, M. E., & Brito, N. H. (2023). OWLET: An automated, open-source method for infant gaze tracking using smartphone and webcam recordings. Behavior Research Methods, 55(6), 3149–3163. https://doi.org/10.3758/s13428-022-01962-w

The ManyBabies Consortium. (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. Advances in Methods and Practices in Psychological Science, 3(1), 24–52. https://doi.org/10.1177/2515245919900809