

# A developmental model of multimodal joint attention through cross facilitative learning based on mutual exclusiveness

Tsukasa NAKANO †  
[tsukasa.nakano@ams.eng.osaka-u.ac.jp](mailto:tsukasa.nakano@ams.eng.osaka-u.ac.jp)

Yuichiro YOSHIKAWA ‡  
[yoshikawa@jeap.org](mailto:yoshikawa@jeap.org)

Hidenobu SUMIOKA ‡†  
[hidenobu.sumioka@ams.eng.osaka-u.ac.jp](mailto:hidenobu.sumioka@ams.eng.osaka-u.ac.jp)

Minoru ASADA ‡†  
[asada@jeap.org](mailto:asada@jeap.org)

† *Osaka University, Yamada-oka 2-1, Suita, Osaka 565-0871, Japan*

‡ *JSTERATO Asada Project, Yamada-oka 2-1, Suita, Osaka 565-0871, Japan*

Joint attention is defined as a behavior to attend to the same object that a communication partner is attending to, and can be performed based on the multimodal observation of the partner's behaviors for attending such as looking at or talking about the object. This is an important behavior for social robots as well as humans. However, it is not trivial for the designer to pre-determine when and which modalities should be used to perform joint attention since it depends on the context of communication. On the other hand, it has been reported that 18-months-old human infants can refer the gaze of their caregivers when they learn names of objects from the caregiver (Baldwin et al., 1991). This fact implies that infants around this age can utilize the information of a modality of gaze for learning joint attention through the modality of label. However, the developmental process of such a mechanism of multimodal joint attention has been still unclear. In this study, therefore, we propose a simultaneous learning mechanism of joint attention with other's gaze direction and labels that the caregiver uses. By doing so, we aim not only at building a robot that autonomously acquires multimodal joint attention but also at computationally understanding the developmental process of joint attention in human infants.

In the previous work under such an approach, it has been shown that a robot can acquire joint attention with a modality of gaze simply by learning a mapping between visual patterns of caregiver's face and its own posture when it finds any salient object (Nagai et al., 2003). However, they did not suppose the situation where the caregiver utters labels of objects attended by him/her. In other words, the issue of joint attention with other modalities of labeling has not been addressed. Therefore, we consider a model of learning mappings for joint attention with both modalities of gazing and labeling. To consider more than two mapping for joint attention, we need a method to determine when and which mapping should be used for performing joint attention and obtaining experiences to learn joint attention. In this study, we propose a measure of mutual exclusiveness of each mapping to determine when each mapping should be adopted for performing and learning of joint attention. We expect that the learning of both mappings are facilitated one another since the proposed measure of mutual exclusiveness indicates the goodness of the mappings in this case. After we report the experimental results in the computer simulation of such a cross facilitation of each mapping, we argue that the phenomenon appeared in the simulation where the mapping for joint attention with label facilitates the learning of another with gaze.

## References:

- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 63, 875-890.
- Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). A constructive model for the development of joint attention. *Connection Science*, 15, 211-229.