

A Computational Model of Acoustic Packaging

Lars Schillingmann, Britta Wrede, and Katharina J. Rohlfing

In order to learn and interact with humans, robots need to understand actions and make use of language in social interactions. The use of language for the learning of actions has been emphasized by Hirsh-Pasek and Golinkoff (MIT Press, 1996), introducing the idea of acoustic packaging . Accordingly, it has been suggested that acoustic information, typically in the form of narration, overlaps with action sequences and provides infants with a bottom-up guide to attend to relevant parts and to find structure within them. In this article, we present a computational model of the multimodal interplay of action and language in tutoring situations. For our purpose, we understand events as temporal intervals, which have to be segmented in both, the visual and the acoustic modality. Our acoustic packaging algorithm merges the segments from both modalities based on temporal overlap. First evaluation results show that acoustic packaging can provide a meaningful segmentation of action demonstration within tutoring behavior. We discuss our findings with regard to a meaningful action segmentation. Based on our future vision of acoustic packaging we point out a roadmap describing the further development of acoustic packaging and interactive scenarios it is employed in.