The use of synchrony in parent-child interaction can be measured on a signal level

Matthias Rolf[1] / Marc Hanheide[2, 1] / Katharina J. Rohlfing[3, 1]
Bielefeld University, [1]CoR-Lab, [2]Applied Informatics, [3]Emergentist Semantics Group

Aim

In our approach, we aim at an objective measurement of synchrony in multimodal behavior. The use of signal correlation provides a well formalized method that yields gradual information about the degree of synchrony. For our analysis, we used and extended an algorithm proposed by Hershey & Movellan (2000) that correlates single-pixel values of a video signal with the loudness of the corresponding audio track over time. The results of all pixels are integrated over the video to achieve a scalar estimate of synchrony.

Motivation

In their work, Gogate and her colleagues (2000) revealed that when showing new objects to their children, parents tend to move it synchronously when providing the new label for it. It has been argued that such synchrony can be used to guide infants' attention towards important features of the situation, i.e. the object and the label (Bahrick et al., 2004). Furthermore, it has been proposed that sensitivity towards amodal properties on the learner's side and the modified behavior that provides lots of amodal overlap between modalities on the tutor's side, is a crucial part in social learning scenarios (Zukow-Goldring, 2006; Rolf et al., to appear). However, so far, in relevant research, subjective coding procedure was applied. In our analysis, we focus on the question whether there is more signal-level synchrony in child-directed interaction.

Synchrony and Correlation

Research on inter-modal perception of synchrony mostly builds on the notion of events (Gogate & Bahrick, 2001; Matatyaho et al., 2007; Virsu et al., 2008). Yet a closer look reveals that most studies lack a precise definition of "synchrony", and a common formalization of this term or the related concept of an event is missing (Hershey & Movellan, 2000). However, we must consider any visual and auditory perception to be faced with a non-structured stream of stimuli. Synchrony must be defined in order to end up with a structured representation, instead of requiring it. The challenge is thus to define synchrony measures on a low level and pre-attentive features. At that level, in our approach, we use temporal correlation between signal flows, which can be seen as immediate formalization of synchrony: stimuli gain high correlation when the signal-values decrease or increase simultaneously in several modalities.

Method

Synchrony in Time and Space

We first consider each pixel on its own. Over a small temporal window, the Pearson correlation between a visual and an auditory feature is computed, before the time window is shifted further. The correlation is used to determine the mutual information (MI) between audio and video as measure of synchrony for each pixel. Since correlation does not consider the stimulus-significance, we excluded insignificant stimuli with a two-staged filtering.

Quantitative Synchrony Estimation

In order to end up with a scalar estimate of synchrony for each one video, we first averaged the mutual information across pixels and frames. As baseline, we compute the same measure, but with pure noise in the audio channel. The final measure of synchrony is the ratio between original-sound MI and pure-noise MI (Fig. 1).

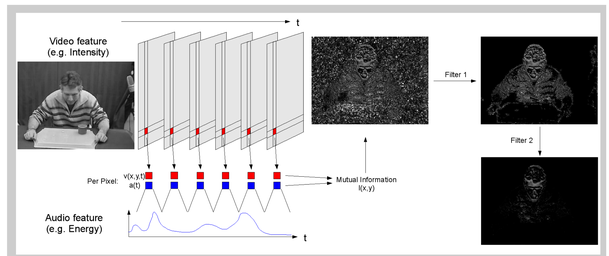

Fig 1. Computational synchrony detection

Setup

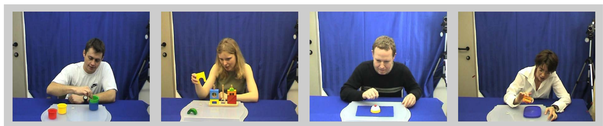

Fig 2. Investigated tasks:
stacking cups, assembling wooden bricks, ringing a bell and using a salt shaker.

Our setting involved 48 parents demonstrated four tasks (Fig. 2) to both their infants and their partners. As not all runs were usable for our experiment, the total number of videos is 184. The infants' age raged from 8 to 30 months. Each parent was instructed to demonstrate the function of the objects to the child (adult-child, AC). Here, the parent was free to teach either the word, the action, or both. The child was attending to the demonstration and interacting with the parent. In a following adult-adult interaction (AA), the same parent was asked to demonstrate the object to her or his partner.

Hypotheses

In agreement with Gogate et al. (2000) we hypothesized that parents produce more synchrony in child-directed interaction in order to arouse and guide the infants' attention, and to structure the interaction. We predicted that more synchrony will be found in AC conditions than in the corresponding AA conditions.

Analysis

Measurement

We measured synchrony for each video separately. The analysis was restricted to segments in that the parent speaks. We used intensity/grey-scale images and gradient images and video features and audio energy (i.e. loudness) for the auditory domain with time windows 1/0.1, 1/0.05 and 1/0.02 frames, with the video sampling rate being 25 Hz.

Evaluation

We initially measured the median synchrony of all adult-adult and of all adult-child interactions, in order to get an overall comparison. We used the immediate pairing of AA/AC conditions to test the Null Hypothesis *H0: P(SyncAA<SyncAC)= P(SyncAC<SyncAA)=0.5* with a two-tailed sign-test. We therefore directly test whether the AC conditions mostly contain more synchrony than the corresponding AA conditions.

Setting		Median		
Feature	Length	AC	AA	Significance
int	0.1	3.48	2.96	0.005
int	0.05	3.86	3.09	0.001
int	0.02	2.73	2.57	–
grad	0.1	2.31	2.00	0.001
grad	0.05	2.68	2.32	0.001
grad	0.02	2.18	1.96	0.1

Tab 1. Median values of synchrony for adult-adult and adult-child conditions, as well as significance against H0 for various settings.

Results

The evaluated medians of synchrony consistently show more synchrony in the AC condition for all feature and time-window combinations. The Null Hypothesis stating that synchrony is equally likely to be higher in AA or AC can be rejected with high significance in most of the settings. Exemplary, the results for gradient images and time window 1/0.05 are plotted below, where each point reflects one pair of AA and AC videos. Despite the variance in the data, even for the subtasks the median synchrony in consistently higher in the AC conditions, indicating a strong effect. Example frames from two videos show that most synchrony is indeed often found on a shown object, which confirms the guidance aspect and goes in line with Gogate et al. (2000). In contrast, a standard saliency model is distracted in many situations.

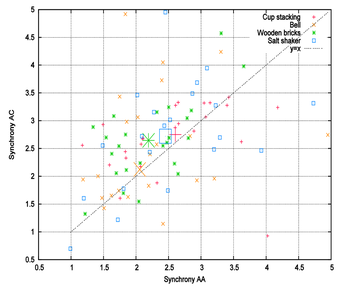

Fig 3. Example with positions of the highest audio-visual synchrony (red) and highest visual saliency [Itti, Koch 1998] (green)

Conclusions

We presented a model to detect the on- and offsets of events implicitly present in the perceptual signals by synchrony. Analyzing persons' behavior in an adult-child and adult-adult interaction, we found more synchrony, i.e. „multimodal motherese" (Gogate et al.. 2000: 219), when parents interacted with their children. Thus, by means of the implicit event boundaries, information seems to be structured and learning relevant information seems to be conveyed by parents, who ostensiveley teach their children actions and objects' functions.

References

Gogate, L., Bahrick, L. & Watson (2000): A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development 71,* 878–894.

Gogate, L. & Bahrick, L. (2001): Intersensory redundancy and 7-mont-old infants' memory for arbitrary syllable-object relations. Infancy *2,* 219–231.

Hershey, J. & Movellan, J. (2000): Audio-vision: Using audio-visual synchrony to locate sounds. *Advances in Neural Information Processing Systems 12*, 813–819.

Matatyaho, D., Mason, Z. & Gogate, L. (2007): Word learning by eight-month-old infants: The role of object motion and synchrony. Paper presented at the 7th International Conference on Epigenetic Robotics.

Rolf, M., Hanheide, M. & Rohlfing, K. J. (to appear): Attention via synchrony: Making use of multimodal cues in social learning.

Virsu, V. Oksanen-Hennah, H. Vedenpää, A. Jaatinen, P. & Lahti- Nuuttila, P. (2008): Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. Experimental Brain Research 186, 525–537.

Zukow-Goldring, P. (2006): Assisted imitation: Affordances, effectivities and the mirror system in early language development. In: Arbib, M. A. (Ed): *From action to language.* Cambridge: Cambridge University Press: 469–500.

Itti, L., Koch, C., Niebur, E. (1998): A Model for Saliency-based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Analysis and Machine Intelligence 12(11), 1254–1259.

Acknowledgments

Katharina Rohlfing's research was supported by the Dilthey Fellowship (Volkswagen Foundation) and by the European Community under the Innovation and Communication Technologies programme of the 7th Framework for ITALK-project (ICT-214668).

 contact: mrolf@corlab.uni-bielefeld.de