Visual attention during linguistic-visual conflicts

Brian Sullivan[1], Harm Brouwer[2], Alba Rodriguez[3], Ricardo Matos[1],
Tommy Strandvall[1], Matthew Crocker[2], Pia Knoeferle[3]
1.Tobii Technology AB, 2. Saarland University, 3. Bielefeld University

Motivation

Eye movements can be directed by visual and linguistic cues. What occurs when this information conflicts in category or timing?

Knoeferle and Crocker (2006) proposed the coordinated interplay account (CIA) of sentence comprehension and linguistically mediated visual attention. The CIA has further refined as an explicit computational model (Mayberry et at., 2009).

Knoeferle, Urbach, & Kutas (2014) incorporated conflicts at the sentence level into the CIA, predicting longer RTs when a visual scene and accompanying sentence mismatched. We wish to extend the CIA as object label agreement and temporal coordination of visuo-linguistic information appear to be important dimensions of comprehension.

Are visual and verbal category labels integrated, independently? Do conflicts affect saccades?

To inform the model, we created a decision making experiment, the ***'Visual-Linguistic' Sorting Game.***

Auditory and visual information manipulated:

1) Matching or conflicting category information
2) Varying stimulus onset asynchronies (SOAs)

If conflict and stimulus timing are independently processed, we should see additive effects of these two factors. Alternatively, if these dimensions interact, they should elicit reliable interaction effects in the statistical analyses.

This is first in a series of experiments to test different components of visuo-linguistic processing related to visual attention (see Future Controls & Comparisons).

Future Controls & Comparisons:

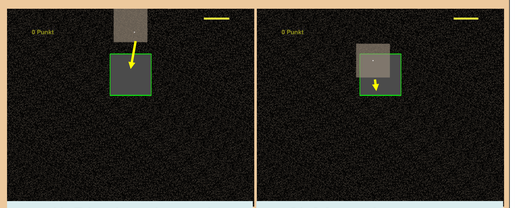
Simple Visual-Verbal conflict RT
- What cost comes from conflicts without eye movements?

Two Category 'Visual-Audio' Sorting
- Is the current experiment testing language or auditory cue classification?

Abstract Linguistic Sorting
- Sort audio-visual stimuli based on abstract categories (i.e. animate vs. inanimate objects)

The presented work was made possible by the Marie Skłodowska-Curie Initial Training Network LanPercept, financed by the FP7 of the European Commission (contract-N° 316748).

Task



Actual trials lasted up to 7s with no text labels

Methods

We used a 3x2 Design (SOA x Conflict Presence). 132 trials (102 congruent & 30 conflict trials, 44 trials of each SOA (see Category & timing conflict panel)

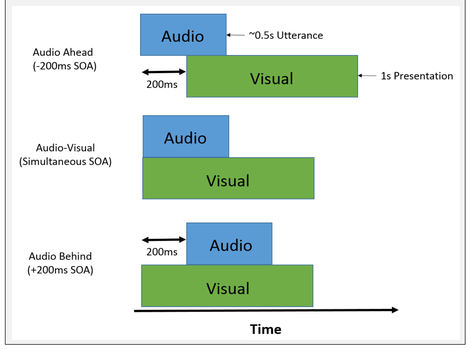
Data collected from 12 native German speakers Participants' eye and mouse movements were tracked during a sorting task (see above)

Each trial, a square appeared onscreen without a category. The participant selected the square and moved it to a 'sensor' to reveal its category. The participant would hear a linguistic category utterance and briefly see the visual category. After which participants sorted the block.

Category & Timing Conflicts

On some trials, visual and auditory information mismatched, e.g. a picture of a monkey was paired with the word 'snail'.

Audio-Visual stimulus onset was also manipulated (+/-200ms).



Fixation Behavior: Single Trial Example

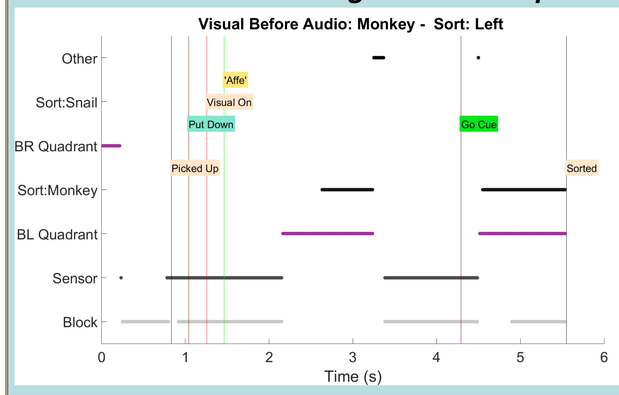


Figure 1: *Timeline of stimulus presentation and labeling of eye tracking AOIs.*

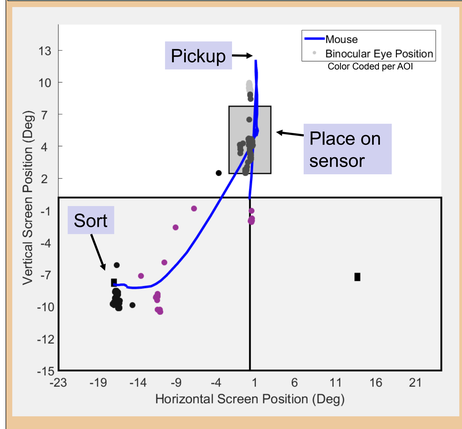


Figure 2: *View of Eye & mouse data for one trial*

Results

Sorting Behavior: Subjects were **96% (+/-4%) correct on congruent trials**. Conflict trials were rewarded at random such that audio or visual cue was right 50% of the time. On conflict trials **54% (+/-24%) of the time the block was sorted in the area related to the audio cue**.

Conflict Trial Behavior

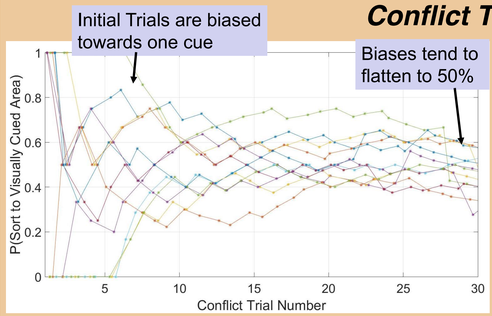


Figure 3: *(Top) Cumulative probability of mouse sorting behavior being directed sorting area indicated by the visual element of the audio-visual cue stimulus*

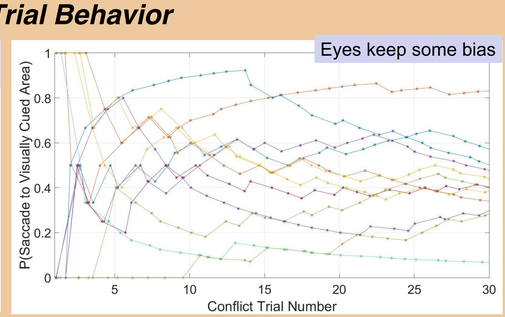


Figure 4: *Cumulative probability of initial saccade after stimulus onset being directed to the sorting area indicated by the visual element of the audio-visual cue stimulus*

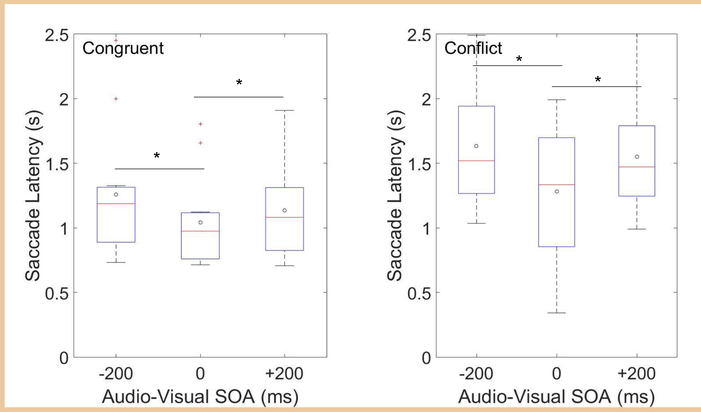


Figure 5: *Latency of Initial Saccade to Sorting Area. Boxplots show median in red, mean as black circle, boxes cover 25th-75th percentile, dashed black lines show range of all data except outliers. (Left) Saccade latencies are depicted for congruent trials and (Right) conflict trials. GLM Repeated measures finds a main effect of both Conflict ($F(1,11)=12.4$, $p<0.01$) and Audio-Visual SOA ($F(2,10)=11.4$, $p<0.01$)*

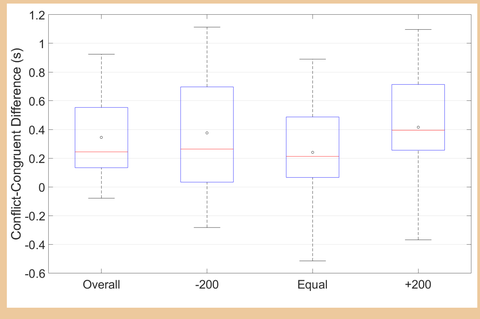


Figure 6: *Within Subjects Saccade Latency Difference between Congruent & Conflict Trials.. On average participants are ~300-400ms slower to initiate a saccade during a conflict trial*

Conclusions

Our novel setup taps into language comprehension as well as decision making, reinforcement learning & multi-modal cue combination. Results show **a main effect of conflict on saccade latency (~200ms slower in conflict trials),** and a main effect of stimulus onset asynchrony. **Mismatches in audio-visual onset latency further slow latencies by ~200ms.**

Results thus far suggest that category and timing conflicts are independent and additive. **See Future Controls & Comparisons panel for plans to test these results.**

The CIA model currently processes words as discrete chunks of equal time duration. **We plan to extend the CIA model to incorporate fine-timing information to explain current results.**