

Pointing Gestures Guide Attention and Ease Load in Simultaneous Interpreting

Objectives Visual context is known to steer comprehension in everyday listening, yet it is unclear whether professional interpreters—who must comprehend and produce speech in parallel—exploit such cues with equal immediacy. We examined whether a pointing gesture speeds reference resolution and reduces cognitive load, and whether a misleading gesture is costlier than having none.

Methodology Twenty-four Geneva-based conference interpreters (22 women, 2 men; mean age 41.7 yrs; mean experience 11.6 yrs) interpreted 30 English videos into French or Spanish while their right eye and pupil were tracked at 1 kHz. Each clip showed a speaker on the left and two pictures on the right. At adjective onset (“large camel”) the speaker pointed to the referent (congruent), to the distractor (incongruent), or remained still (neutral). Fixations to the speaker’s face, hand, target and competitor were analysed for qualifier, noun and spill-over windows with generalised additive mixed models; baseline-corrected pupils indexed effort.

Results Congruent pointing produced an immediate and sustained target advantage: gaze shifted to the correct picture within ≈ 200 ms of gesture onset and stayed there. Incongruent pointing first drew gaze to the distractor, but the error was fully repaired by noun onset. Neutral trials showed no early bias; a target preference emerged only at the noun. Pupils told the same story: during noun and spill-over windows, neutral trials elicited larger dilations than incongruent trials, indicating that searching without guidance is more taxing than correcting a miscue. Gesture type did not modulate pupils during the qualifier.

Discussion A pointing hand functions as a fast, high-priority cue that survives the heavy dual task of simultaneous interpreting. Any spatial anchor narrows the search space, so the cost of revision is smaller than the cost of establishing reference unaided. Conversely, missing gestures inflate effort even when speech itself is unambiguous. Clear, well-timed video feeds are therefore not a luxury but a necessity for interpreters. The study extends laboratory evidence on gesture–speech integration to the extreme time pressure of professional booths and invites future work on more naturalistic discourse and chronically delayed feeds.