Review of "Real-time lexical comprehension in young children learning American Sign Language" (DS-06-17-0234-P)

Summary: The manuscript reports an experiment contrasting the ability of deaf and hearing children learning ASL (as well as proficient deaf adults) to redeploy attention from a signer to a visually present referent as language input is perceived in real time. The results show that all participant groups can shift gaze to intended targets before signs are completed in their entirety, which is taken to reflect incremental recognition of signs. Further, although hearing and deaf learners were relatively slower than adults in the time to reach intended target images, and were less accurate, patterns did not differ for the two child groups. The results also show the speed of sign recognition correlates with vocabulary size, similar to effects found in studies of spoken language.

Assessment: I enjoyed this manuscript. The study seems well designed and the results are fairly straightforward. I believe, however, that aspects of the manuscript would benefit from some reordering and reframing, and that some narrowing in the paper's focus might be beneficial. Comments below.

1. One concern is that methodological and theoretical issues get a bit tangled at various points. This could be readily handled with some rewriting and reorganization. As background, there is (as the authors note) existing evidence that signs are interpreted incrementally by adults (Emmorey & Corina) and that this can be assessed using a modified listening-while-looking methodology, where participants have to shift gaze from the signer to visually co-present referents (Lieberman et al.). So the immediate question is whether the latter methodology can be used with children, and if so, whether learner types differ, and whether children show correlations with vocabulary, as in studies of spoken language. Positive results would indicate that (i) the link between vocabulary size and processing speed is a modality general aspect of language behaviour in children, and (ii) that the methodology could be valuable for future studies of sign language acquisition. Beyond these core points, however, I don't see the rationale for a number of other things that have been brought into the paper. As one example, I don't see the argument that immediate gaze shifts constitute an important skill for children to learn (see p. 5) and I'm not sure the paper actually addresses issues from this angle anyway. For example, if children waited an additional 100 ms or more until the end of a sign before shifting attention to visually-present object, would this be expected to change the course of language acquisition in some meaningful way, or have important implications for nature of lexical or referential processing? Similarly, if children's gaze shifts only occurred after a sign was complete, would researchers really take this as evidence against incremental interpretation? I believe an equally likely conclusion would be that children's ability to redeploy gaze from an attentionally-attractive location in visual space (a moving person) to another location is simply slower than in adults, making the gaze tracking methodology ill-suited for this age group. So the rapid gaze-shifting ability strikes me as really being a methodological issue in relation to what the authors want from

the eye movement measures, not a theoretical one. Another statement that (respectfully) doesn't really seem to fit with the study at hand is on p. 18: "Parallel looking patterns for deaf and hearing ASL learners suggest that both groups are sensitive to modality-specific constraints of processing a sign language". Is this really what the results tell us, and what constraints are at issue here? Linguistic reference to things in the here-and-now (which sets up the relevant visual competition in the current case) instead of "absent" referents doesn't strike me as a defining feature of sign language use. Further, even in face-to-face spoken communication, past research tells us that listeners like to look at talkers' faces (although this is clearly not how spoken language eye tracking studies are typically conducted). I'm afraid I don't see the modality-specific argument.

One suggestion would be to stick to a slightly more streamlined set of issues, and to reorder things so that the question of whether children can shift gaze before a sign is completed sits in the #1 spot (instead of #3). Once this point is firmly established, I think the remaining points and their corresponding measures will have a more logical place in the flow of the argument, e.g., now that we know the gaze shifts are occurring as signs unfold in time, it makes sense to explore differences in this real time processing across participant groups and potential correlations with vocabulary size. Indeed, on pp. 19-20 the authors themselves make the point that their finding that participants can quickly shift attention away from the speaker validates the inference that timing measures can reflect the speed of lexical access. So I do think there would be value in reordering the way in which things are presented to make the value and role of the different measures more apparent.

2. The introductory section might benefit from some more consideration of when eye movements reflect lexical vs. simply referential processing. This feels a bit fuzzy at times. The challenge, of course, is that eye movement paradigms make use of an overt referential behavior (linking an auditory or visual symbol to a real-world referent) as a measure for making inferences about lexical processing. But I think a bit of rewriting can sharpen things up.

3. The motivation for the comparison between hearing and deaf native signers could be clearer. I think there is definitely value to this comparison, but the motivation as currently stated comes across a bit weak. The authors speculate that there may be differences based on the groups' differential access to auditory information. But why exactly would this matter when performing what is essentially a purely visual task? Is the specific idea that the hearing children could be more susceptible to auditory attentional capture (assuming the testing environment, like their home environment, is not totally free from environmental noise)? Or could differences arise because hearing children might be "splitting their time" across ASL and English and as such be less fluent compared to monolingual ASL learners even when matched on certain measures? Also on this point, I think speculations for possible differences for the two younger groups may need to be discussed separately when it comes to the question of whether gaze can be rapidly disengaged from the speaker vs. the question of whether signs are processed

rapidly/incrementally. Measurement issues aside, one could imagine, for instance, that native deaf signers might be worse at the first but better at the second. [Related to this: p. 23:19 "…while the hearing children could use vision and hearing to process incoming information." This gives the impression that there was speech information in the current stimuli. There wasn't, was there?)

4. There were a few methodological details that were not clear to me. Did each participant encounter 8 trials in total? Were they balanced for the two different talkers and the two different question types, or were these cycled across different subsets of participants? (Did every participant see exactly the same stimuli?) Also, although the authors note that the item pairs have "minimal phonological overlap" (p. 10), I think readers will want to know about the precise extent to which the various target and competitor signs were distinct with respect to ASL parameters like handshape, orientation, location, etc. There were also some other places where there could be more clarity in describing the stimuli (e.g., p. 10 line 55: the earlier examples made it sound like the target sentences are in fact questions. Was there an additional question after those questions?)

5. One methodologically and theoretically important issue I would like to see addressed concerns the issue of iconicity in the signs used as stimuli. The sign for ball, for example, seems to have a clearly iconic aspect to it, reflecting shape characteristics that can be easily mapped to the target object on the basis of low-level attentional features. Was this true of other signs as well? This seems important because iconic signs could plausibly exaggerate the speed and/or accuracy of processing. If measures of iconicity are available for ASL signs, one possibility would be to include iconicity as an item-wise predictor in the statistical analyses.

6. Re: measures of interest. The operationalization of "sign onset" seems quite different from "word onset" in studies of spoken word recognition (and is perhaps closer to 'uniqueness point' than word onset?). The initial speech sound in a word is rarely, for example, sufficient to entail 100% agreement across listeners regarding target identity, in part because of the nonlinear nature of speech cues (e.g., with information about the place-of-articulation of consonants being carried by vowel formant transitions, etc.). I understand the challenges here ,but it seems relevant to consider whether the sign onset measure as implemented might exaggerate measures of incremental interpretation (proportion of target processed), or alternatively whether "incremental" (rather than "rapid", etc.) is really the right term to describe the processing of information that is already fully available at the left margin of the measurement interval. Also: how exactly was this sign onset measure calculated? Were the adult signers watching videos in slow motion, or via a gating task?

p 14, sentence beginning "In studies with adults…". This sentence might need a tweak or two, as the truth of this statement really depend on the design. If for example there is a phonological competitor in the display (e.g., candy + candle), then

initial fixations can simply reflect early lexical hypotheses, not the speed of lexical access for the intended target.

It would be good to know the rationale for using an analysis window for the accuracy measure that ranges from 600-2500 ms after the onset point.

The authors report separate statistical models using age vs. vocabulary predictors is because this predictors are highly correlated. I think it likely makes sense to report this correlation somewhere (maybe I missed it).

7. Re: Discussion section.

i. Bottom of p. 25 and highlights: "…as soon as listeners have enough information …". Is this perhaps a bit too strong, given that "sign onset" was identified as the point in the video at where there was sufficient information for target identification yet in absolute terms gaze shifts tended to occur toward sign offset?

ii. How do the results for adults compare to the Lieberman et al. study cited in the introduction? (p. 4:36)

iii. I would urge the authors to be just a bit more cautious when describing the age-related differences in reaction time in the discussion. This is because there was no control condition that could assess the possibility of age-related gains in the efficiency of controlling visual attention to scene regions.

8. MINOR POINTS

p. 3:27. Introduction: The description of adult patterns in studies of spoken word recognition (i.e., "as soon as the auditory information is sufficient") sounds as if it applies to infants, whereas in reality infants are measurably slower and show delays in target identification in relation to the uniqueness point within a word (as the authors note elsewhere).

p. 4:19. (Sentence beginning "As in spoken language…"). Most current models of spoken word recognition avoid talk of 'stages' and instead favor the notion of cascaded processing. It may be helpful for the authors to clarify what they mean here (or just eliminate this wording if not needed to make a specific point).

p. 4:57: "…less of the linguistic signal" Please clarify for the readers (e.g., does this mean a shorter time sample of the unfolding linguistic signal?) Also, the description of the different gating tasks could be clearer for people unfamiliar with this methodology. As it stands, "increasingly longer segments" risks sounding like the participant hears a word or sees a sign in its entirety several times, but each time more slowly / stretched out.

p. 5:52: The allocation of visual attention is not a 'basic learning mechanism'. Please clarify what is intended here.

References section: A Lieberman et al. study is listed twice in the references, with two different dates. A (different?) Lieberman et al. study mentioned in the introduction seems to be missing. Also, some references do not conform to APA style (e.g., handbook entries, capitalization rules), and some references run together in a few places.