Examining the spoken language input to infants with cochlear implants

Lillianna Righter*[1], Alex Emmert*[1, 2], Erin Campbell[3], Derek Houston[4], & Elika Bergelson[1]

[1] Department of Psychology, Harvard University, Cambridge, MA

[2] Department of Linguistics, University of Maryland, College Park, MD

[3] Deaf Center, Boston University, Boston, MA

[4] Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT

Author Note

*AE & LR co-first author.

Correspondence concerning this article should be addressed to Lillianna Righter*, 33 Kirkland St, Cambridge, MA 02138. E-mail: larighter@fas.harvard.edu

Abstract

We examine the spoken language environments of 16 deaf and hard of hearing infants with cochlear implants (DHH), 16 hearing chronological age matches (CAM), and 16 hearing age matches (HAM), ages 14-32 months. Using manual annotations and automated LENA analyses (Xu, Yapanel, & Gray, 2009), we find overall similarities in quantity of language input and the social, linguistic, conceptual, and auditory features of the language environment of each group. Caregivers use slightly longer MLU to hearing children, and use more highly auditory-associated words with DHH children. We find differences in children's vocalizations and conversational turn count, with DHH children producing fewer and less mature vocalizations and engagin in fewer conversational turns. These findings replicate prior literature and suggest that caregivers do not adapt their speech on the basis of infants' perceptual capacity. However, they likewise reinforce prior findings that the amount of linguistic input and interaction they receive is shaped by infants' own language productions. This suggests any DHH infants' difficulties with productive language outcomes 1) is not fundamentally due to differences in language input behavior from caregivers; but 2) may slow the quantity of language input and interaction that they receive. Instead, differences in language outcomes between hearing and DHH children may be driven by decreased access to and difficulty processing the language environment because of the noisy signal from a cochlear implant. Full paper forthcoming.

Examining the spoken language input to infants with cochlear implants

```
## [1] "/Users/bergelsonlab/Desktop/git/DHH_public_code_sample"
```

Methods

All interaction with human subjects, data collection, and storage procedures were conducted in accordance with the guides laid out in the Declaration of Helsinki. All activities were approved by Institutional Review Boards at Duke University, the Ohio State University, or Harvard University.

Participants

A total of 16 deaf/hard-of-hearing (DHH) children with cochlear implants contributed recordings during a larger study conducted at the Ohio State University (for general results about the larger sample, see Wang, Cooke, Reed, Dilley, & Houston, 2022). All DHH children in this sample experience bilateral severe-to-profound hearing loss, use bilateral cochlear implants (age at first activation 7.96-23 months, M = 13.92 months), and are acquiring spoken English as the target language; minimal to no sign language exposure was reported.

Table 1

Demographic information by group, n = 16 per group.

CAM	CI	HAM
Age		
M = 21.1 mo., 14.06 to 32 mo.	M = 20.7 mo., 14.03 to 32 mo.	M = 6.9 mo., 6 to 9 mo.
Hearing Age		
M = 21.1 mo., 14.06 to 32 mo.	M = 6.8 mo., 5.6 to 9 mo.	M = 6.8 mo., 6 to 9 mo.
Age at first activation		
	M = 13.9 mo., 7.96 to 23 mo.	
Gender		
Female 81.25%	Female 81.25%	Female 81.25%
Male 18.75%	Male 18.75%	Male 18.75%
Maternal education level		
Less than high school 6.25%	Less than high school 12.5%	High school diploma 18.75%
High school diploma 12.5%	High school diploma 25%	Some college 25%
Some college 31.25%	Associate's degree 12.5%	Bachelor's degree 37.5%
Associate's degree 6.25%	Bachelor's degree 31.25%	Advanced degree 18.75%
Professional certification 12.5%	Advanced degree 18.75%	
Bachelor's degree 18.75%		
Advanced degree 12.5%		
Race		
White 62.5%	White 87.5%	White 87.5%
Multiple races 6.25%	Multiple races 6.25%	Multiple races 6.25%
Unreported 31.25%	Unreported 6.25%	Unreported 6.25%
Ethnicity		
Not Hispanic or Latino 68.75%	Hispanic or Latino 6.25%	Not Hispanic or Latino 93.75%
Unreported 31.25%	Not Hispanic or Latino 87.5%	Unreported 6.25%
	Unreported 6.25%	

Each DHH child was matched with two typically-hearing children: one based on chronological age (CA) and one based on hearing age (HA). Hearing age was operationalized as the amount of time that children had auditory access to spoken English. For typically-hearing children, this is the same as chronological age, as they have had access to sound from birth. For DHH children, this is the amount of time since activation of their first CI, or in other words their age at the time of recording minus their age at activation. As a result, HA matches are younger than DHH children and their CA matches by design.

Recordings from typically-hearing children were gathered from preexisting English-speaking corpora or collected from the Durham, North Carolina area. Typically-hearing children were monolingual English learners (parents reported that at least 75% of children's language input was spoken English), and were matched to DHH children based on infant sex, maternal education (within one level), and number of older siblings (none, one, two, or 3 or more; twins were matched to twins). The age matching guidelines were based on the age or hearing age being matched: under twelve months, the control infant's age was within +/- 2 weeks; between 12 and 24 months the control infant's age fell within +/- 1 month of the DHH child's age; and over 24 months, the infant's age was within +/- 2 months difference. The full distribution of demographic factors can be found in Table 1.

Data collection

Each child contributed one day-long recording (48 recordings, mean duration = 14.37 hours) using LENA devices (Ganek & Eriks-Brophy, 2016; Gilkerson & Richards, 2008; Zimmerman et al., 2009). Parents were instructed to start the recording when the child woke up and to keep it nearby when the vest had to be removed (e.g. for baths or naps). Parents received instructions for pausing and resuming recordings in the case of private conversations, and were given the option to have any part of the recording deleted after data collection and not analyzed, if they chose.

Data analysis

Each recording was algorithmically analyzed in its entirety by LENA software (Xu et al., 2009), and a portion of each recording was further transcribed and annotated by trained annotators using ELAN (versions 5.7- 6.8; (Brugman & Russel, 2009; Sloetjes & Wittenburg, 2008)). Fifteen two-minute intervals were extracted randomly in each recording for hand annotation, in addition to five "high-volubility" two-minute intervals containing dense speech, as identified by the voice type classifier for child-centered daylong recordings (Lavechin, Bousbib, Bredin, Dupoux, & Cristia, n.d.). This resulted in 20 two-minute intervals, for a total manual annotation time of 40 minutes per recording. Annotators listened to, but did not annotate, the preceding two minutes and following one minute of each segment's audio to establish context.

Manual annotation was performed in accordance with the ACLEW annotation scheme (Soderstrom et al., 2021), with speech by individuals other than the target child transcribed using the minCHAT transcription style (MacWhinney, 2019). Each non-target-child utterance was classified based on the role of the addressee (child, adult, both child and adult, pet, other, or unknown) and lexically transcribed. The target child's vocalizations were annotated for maturity (non-canonical babbling, canonical babbling, laughter, crying) and lexicality (contains words, single- or multi-word utterance). 30 annotators contributed to this data set over 6 years. We conducted a 10% recode on the closed-set coding categories to assess inter-coder reliability; agreement was 90.6%, Cohen's kappa 0.88, indicating high consistency.

To maximize statistical power given our relatively small sample, we combine the two hearing groups into a single comparison group—unless the two hearing groups differ across age for a given variable. First, we check within the typically-hearing group whether the input variable differs as a function of age, to establish whether we should expect an effect of age on the input variable. That guides our choice of test: if the variable differs across age in typically-hearing children, we run a linear model testing whether the input variable differs by child hearing status while controlling for age (Input Variable ~ Group~Cochlear Implant vs. Typically-Hearing~ + Age). If the input variable *does not* differ across age in typically-hearing children, to conserve power, we combine the hearing age match and chronological age match groups and run a t-test comparing the input variable by hearing status (Input Variable ~ Group~Cochlear Implant vs. Typically-Hearing~).This approach allows us to simplify to a two-group comparison when possible, while preserving the careful demographic matching of both hearing age and chronological age.

Automated LENA Measures. The LENA software generated values for Adult Word Counts (AWC) and Conversational Turn Count (CTC) for each recording. AWC estimates the number of words produced by adults around the child, and defines a conversational turn as a pair of utterances produced by an adult speaker and a child speaker (or vice versa) occurring within within 5 seconds of each other. We normalized both of these measures to a per-hour value based on each recording's length.

We used LENA software further calculated the proportions of Nonspeech Noise, Overlapping Sound, and TV/Media Noise in each recording, expressed here as a simple fraction of each recording that was identified as containing each type of noise.

We first check to see whether the LENA metrics vary across age among typically-hearing participants. Adult Word Count did not vary across age (r = 0.13, p = .484), but conversational turn count (r = 0.75, $p < .001$) and child vocalization count (r = 0.7, $p < .001$) did. For adult word count, therefore, we combined the two typically-hearing groups and compared them to the CI group. Results of the Wilcoxon test showed no significant difference in overall word count between the cochlear implant group and their typically-hearing matches ($\mathrm{Mean}_{\mathrm{CI}}$=1072.58, $\mathrm{Mean}_{\mathrm{Hearing}}$=1254.24, W = 177, p = .086). For conversational turn count, we found that while conversational turn count increases across age for the typically-hearing participants, it did not increase across age for children with cochlear implants (Model $R^2 = 0.41$, $p < .001$, $\text{Beta}_{\text{HearingStatus}} = -37.5$, $p = .061$, $\text{Beta}_{\text{Age}} = -0.26$, $p = .768$, Interaction: $\text{Beta}_{\text{HearingStatus:Age}} = 2.45$, $p = .013$). Results were similar for child vocalization count: while child vocalization count increases across age for typically-hearing children, it did not for children with cochlear implants (Model $R^2 = 0.36$, $p < .001$, $\text{Beta}_{\text{HearingStatus}} = -208.77$, $p = .011$, $\text{Beta}_{\text{Age}} = -3.46$, $p = .323$, Interaction: $\text{Beta}_{\text{HearingStatus:Age}} = 10.9$, $p = .007$).

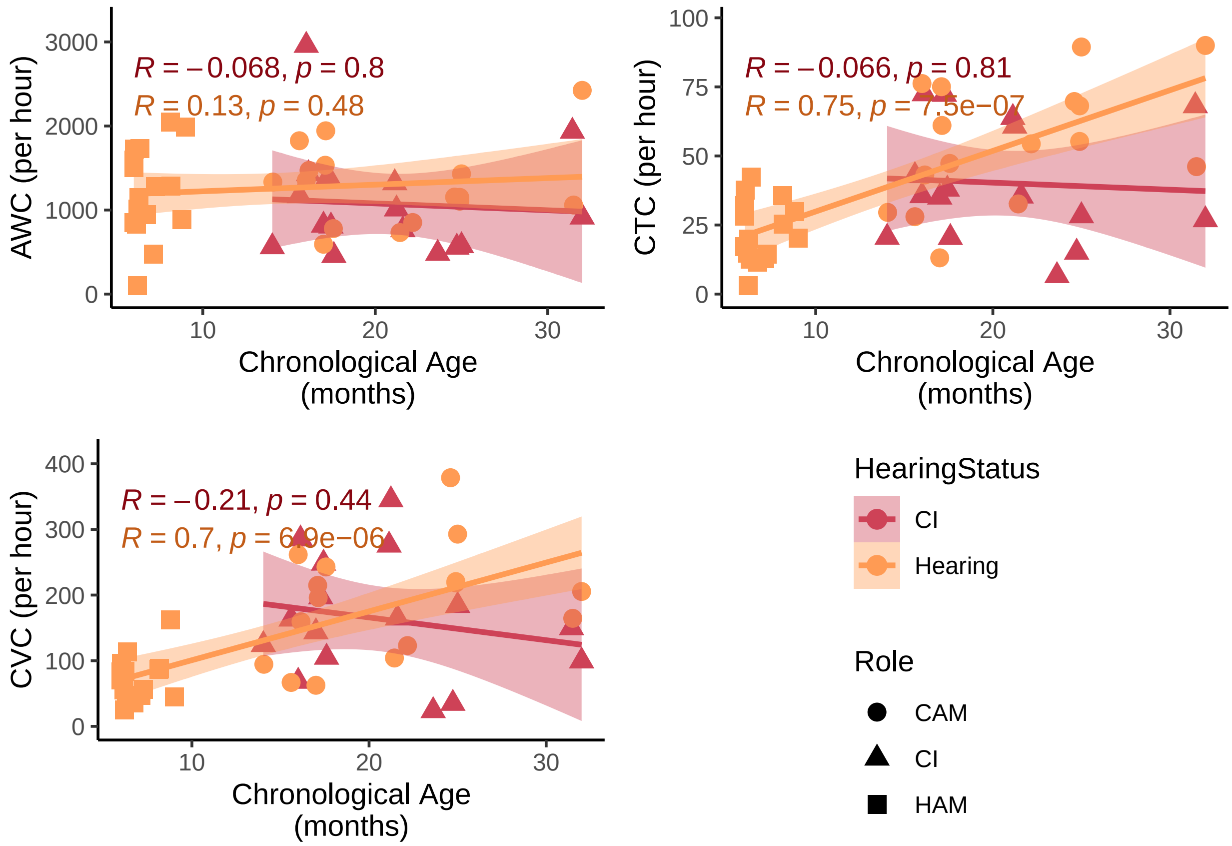

Figure 1. Measures from Automated LENA analysis.

Language Exposure Measures. Total Word Count based on the manual annotations for each recording. This value is a count of all individual words produced by speakers other than the target child. Words were defined as strings separated by spaces in the transcription.

Each manually-coded utterance was annotated for its addressee: child, adult, both children and adults, a pet, other (e.g., virtual assistants, higher powers, themselves), or unknown addressee. While the annotation scheme does not distinguish speech directed to the target child from that directed to other children, looking at speech directed to adults, pets, and others allows for an estimation of the proportion of *types* of speech each group is exposed to, (i.e., child directed speech vs. overheard speech). We calculated the overall proportion of speech directed to each category.

To more closely estimate the overall quantity, we calculated these measures only in the 30 randomly-sampled minutes of transcription, not the 10 minutes selected for high-density talk.

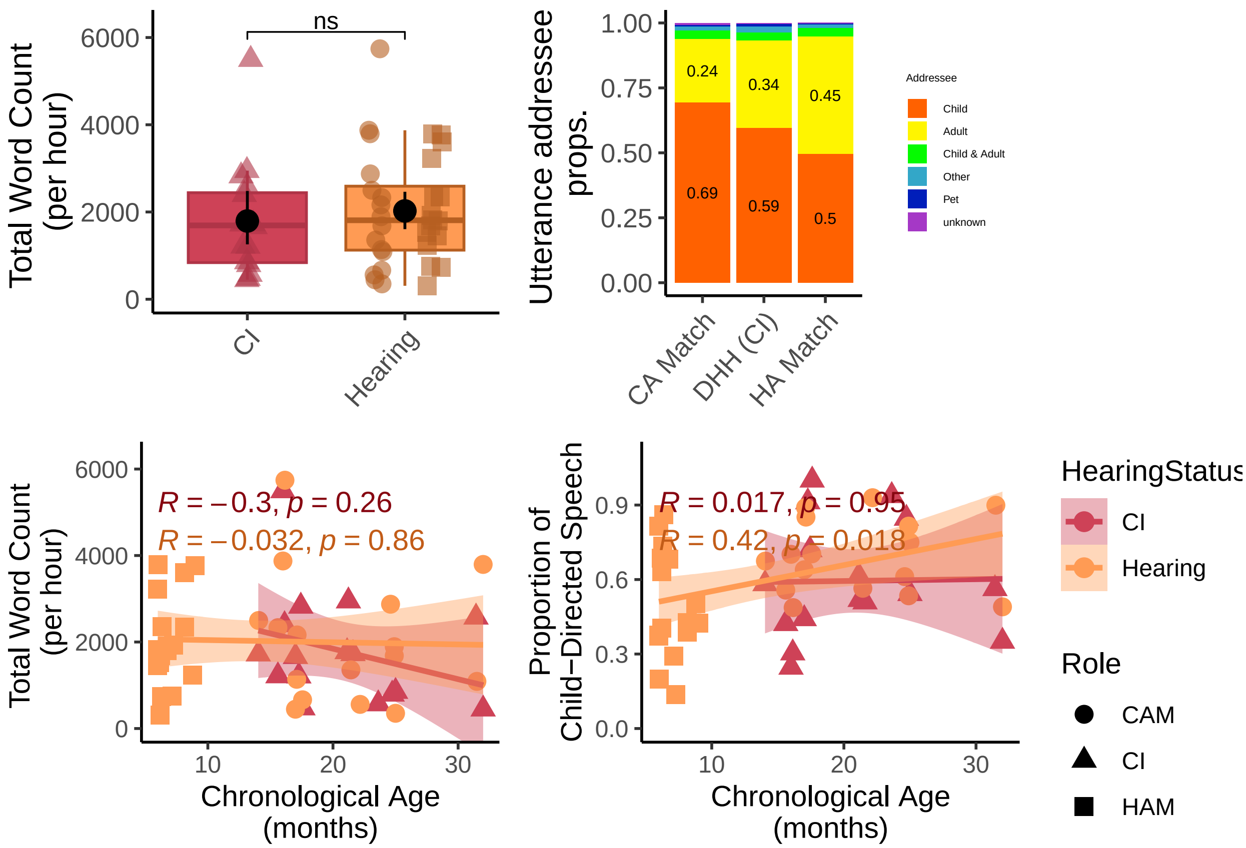

Figure 2. Language exposure measures.

Language exposure was broadly quite similar between the CI group and each hearing match group. Because manual word count did not vary across age for typically-hearing participants (r = -0.03, p = .861), we collapsed the CAM and HAM groups for the word count analysis. Results of the Wilcoxon test showed no significant difference in overall word count between the cochlear implant group and their typically-hearing matches ($\mathrm{Mean}_{\mathrm{CI}}$=1793.38, $\mathrm{Mean}_{\mathrm{Hearing}}$=2024.81, W = 227, p = .537).

For the proportion of child-directed speech, we observed a significant correlation with age among the typically-hearing participants, so we ran a linear model with age and group as predictors (r = 0.42, p = .018). This model does not significantly explain the variance in the proportion of child-directed speech (Model $R^2 = 0.11$, p = .163, $\mathrm{Beta}_{\mathrm{HearingStatus}}$ = -0.13, p = .558, $\mathrm{Beta}_{\mathrm{Age}}$ = 0, p = .942, Interaction: $\mathrm{Beta}_{\mathrm{HearingStatus:Age}}$ = 0.01, p = .371). Based on visual inspection of 2, it seems like the proportion of child-directed speech might increase for typically-hearing children but not DHH children, but as seen in the graph, the proportion of child-directed speech shows wide individual variability, and our analysis does not yield any conclusions. statistical comparisons of other addressee proportions were not performed, as child-directed speech was the primary variable being investigated.

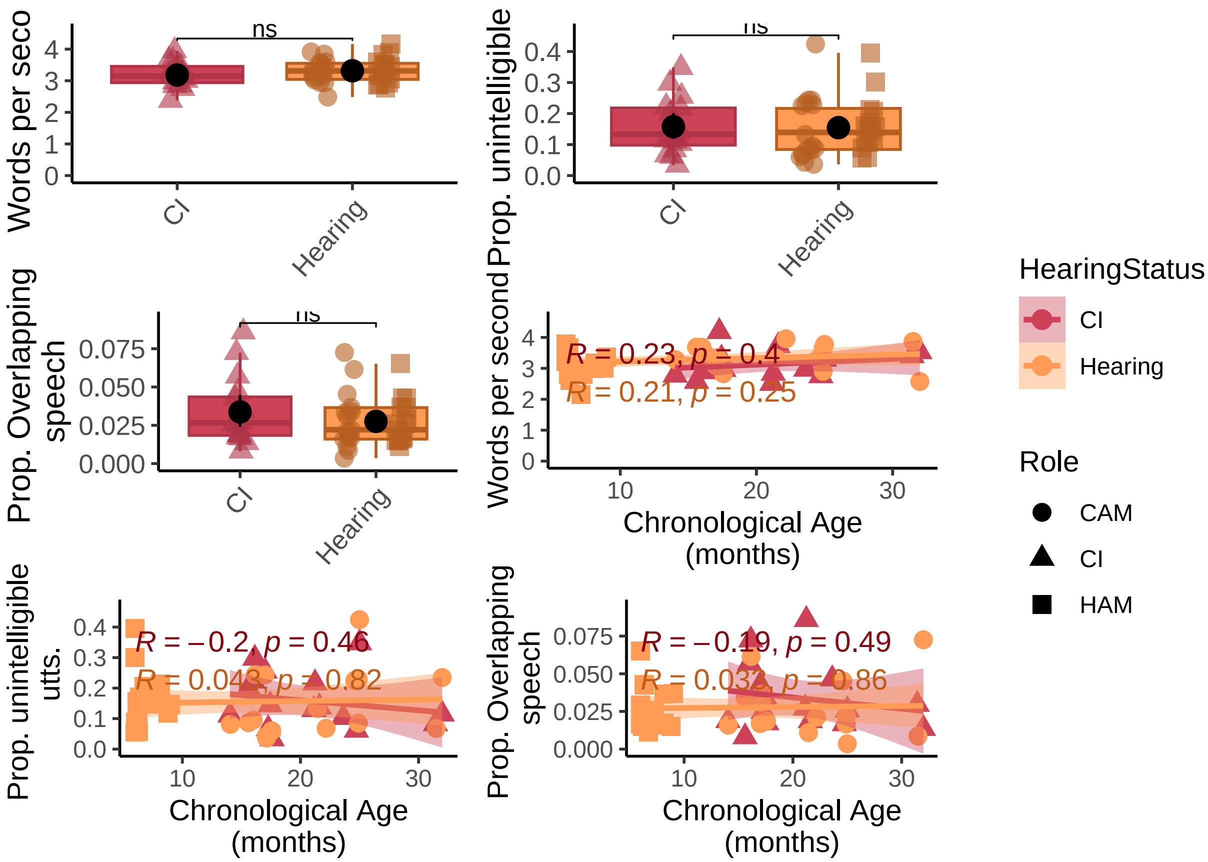

Figure 3. Audibility measures.

Audibility Measures. In addition to automated audibility analyses, we computed three measures of input audibility based on the manual annotations. First, "Words per Second" measures the average rate of speech in the child's auditory environment. For each utterance, the number of words was divided by the duration of the utterance in seconds. These values were then averaged across all of the utterances in each recording. Utterances containing unintelligible speech were excluded from this calculation.

Second, we calculated the proportion of utterances containing speech deemed unintelligible. We note that this measure relied on the determination of intelligibility by an adult, typically-hearing listener listening to a recording and is thus an imperfect (though potentially still useful) proxy. That is, whether speech was or wasn't intelligible to the child cannot be captured, and this measure likely differs from the child's experience in several ways. First, though the child wore the recorder, the physical conditions of the the recorder differ from the child's own ears and cochlear implants (e.g. could be muffled by their shirt when the child is being held). Second, for DHH children, we have no indicator of the acoustic quality of each utterance as it was processed through their cochlear implant. This measure is a proxy for identifying utterances that are far away, muffled, rapid, or obscured by competing sound and are more likely to be difficult for a language learner to process.

Finally, we calculated the proportion of overlapping speech in the manual transcription. Each utterance has an onset and offset time. When two or more utterances overlap in time, we count the overlapping duration towards the total amount of overlapping speech in the transcribed regions of the file. We report the proportion here as the summed duration of overlapping speech divided by the length of the recording.

Next, we investigated whether parents of children with cochlear implants might try to make speech more audible by slowing speech down (speech rate), speaking louder or more clearly (proportion of unintellible utterances), or reducing contexts where there are multiple speakers (proportion of overlapping utterances).

Because speech rate did not vary across age for typically-hearing participants (r = -0.09, p = .625), we collapsed the CAM and HAM groups for the speech rate analysis. Results of the Wilcoxon test showed no significant difference in speech rate between the cochlear implant group and their typically-hearing matches ($Mean_{CI}$=3.18, $Mean_{Hearing}$=3.31, W = 210, p = .323). The proportion of unintelligible utterances also did not vary across age for typically-hearing participants (r = 0.04, p = .817), so we again collapsed the CAM and HAM groups for the proportion of unintelligible utterances analysis. Results of the Wilcoxon test showed no significant difference in proportion of unintelligible utterances between the cochlear implant group and their typically-hearing matches ($Mean_{CI}$=0.16, $Mean_{Hearing}$=0.15, W = 263, p = .888).

Finally, since the proportion of overlapping utterances did not vary across age for typically-hearing participants (r = 0.03, p = .857), we ran a Wilcoxon test comparing the amount of overlap in the input to typically-hearing children versus to children with cochlear implants. The two groups did not differ (Mean_{CI}=0.03, $\text{Mean}_{\text{Hearing}}$=0.03, W = 300, p = .345).

Complexity Measures. We calculated Mean Length of Utterance, quantified as the mean number of morphemes per utterance in the speech input. Utterances' morpheme counts were parsed and counted using the `morphemepiece` package in R (Bratt, Harmon, & Learning, 2022). We excluded utterances containing unintelligible speech.

We also calculated Type-Token Ratio to analyze the amount of lexical variety in each child's input. This measure was computed by "chunking" the input speech into 100-word bins across each group, then calculating the proportion of unique words out of the 100 in each bin. These uniqueness values were then averaged to produce a single value for Type-Token Ratio for each recording. Normalizing the denominator allows for a measure of lexical diversity that is less coupled with the raw quantity of speech in the input (Montag, Jones, & Smith, 2018; **campbell2025?**).

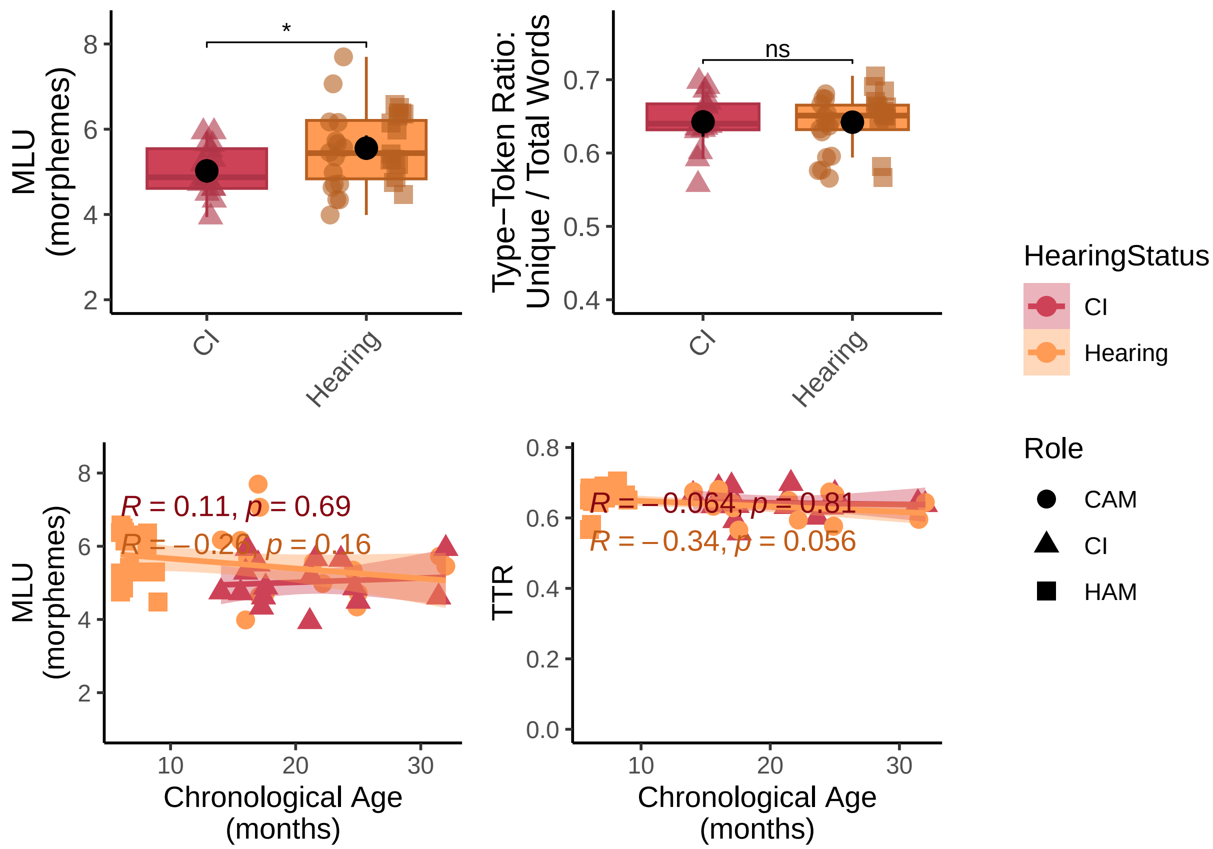

Figure 4. Input complexity measures.

Neither MLU (r = -0.26, p = .157) nor type-token ratio (r = -0.34, p = .056) varied by age in the typically-hearing participants, so for both analyses, we collapsed the two typically-hearing subgroups. We found that MLU was higher for language input to typically-hearing infants (Mean_{CI}=5.02, $\text{Mean}_{\text{Hearing}}$=5.56, W = 160, p = .036). Type-token ratio did not differ by group (Mean_{CI}=0.64, $\text{Mean}_{\text{Hearing}}$=0.64, W = 242, p = .770).

Conceptual Measures. We determined the temporality of each utterance following the procedure in (**campbell2025?**). To calculate this, we used the R package `udpipe` (**wijffels?**) to tag the first verb in each utterance with tense and mood features to determine the temporal quality of each utterance. Past tense, going to/want to/got to, and modal verbs were classified as decontextualized utterances, and present tense and gerunds were classified as present utterances. Fragments and other utterances for which no temporality features were tagged were left unclassified. For more discussion of the benefits and limitations of this analysis, see (**campbell2025?**).

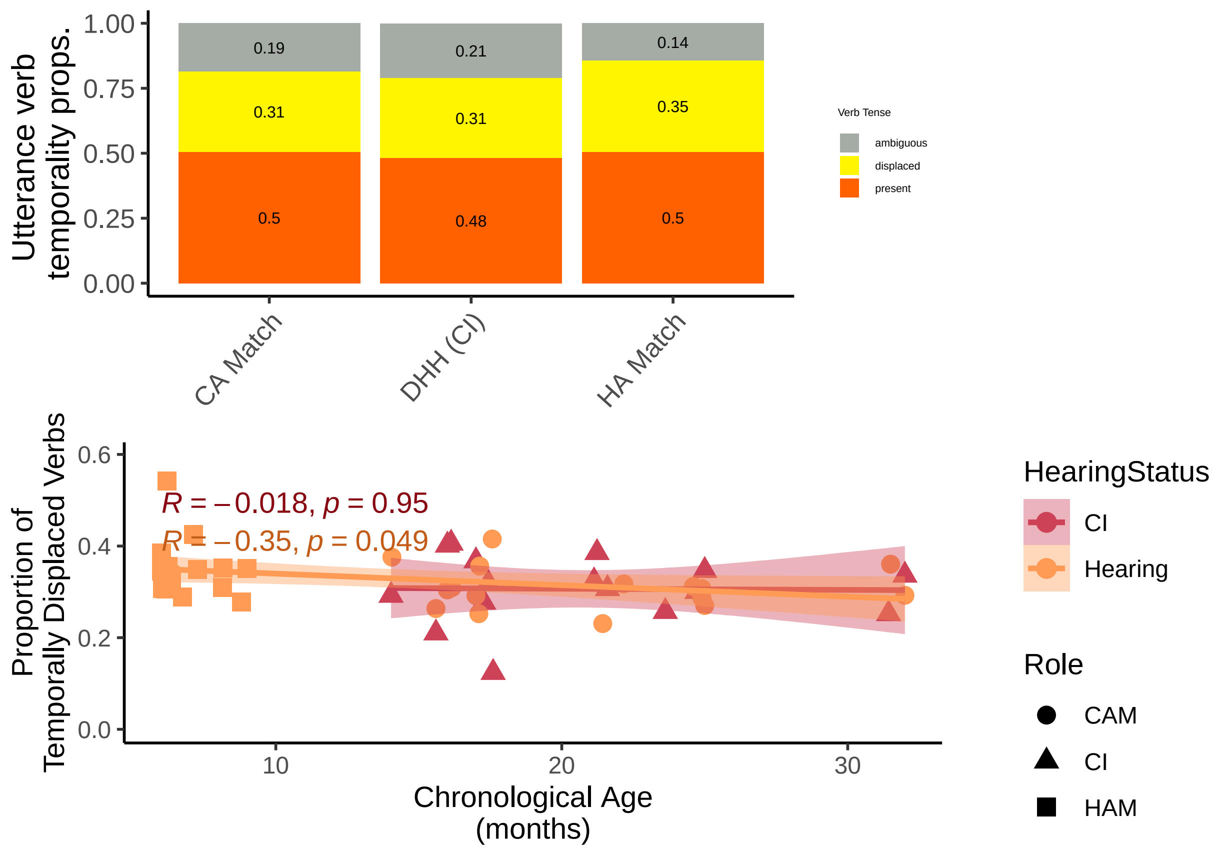

Figure 5. Conceptual Measures.

Verb temporality did not differ across age for our typically-hearing participants (r = -0.13, p = .478), so we collapsed the hearing groups together. Language input to the CI group contained a slightly lower proportion of temporally *present* utterances (Mean_{CI}=0.48, $\text{Mean}_{\text{Hearing}}$=0.5, W = 163, p = .042) but a similar amount of temporally displaced utterances (Mean_{CI}=0.31, $\text{Mean}_{\text{Hearing}}$=0.33, W = 222, p = .468).

Relationship between input and language outcomes. We finally conducted two additional linear models, looking at both automated measures and measures from manually-annotated measures. For automated measures, we examined the effect of AWC on Child Vocalization Count, a numerical estimate expressed by the LENA software of the number of utterances produced by the child. For manual measures, we correlated Manual word Count and the proportion of the target child's utterances that were classified as canonical babbling (which includes lexical utterances).

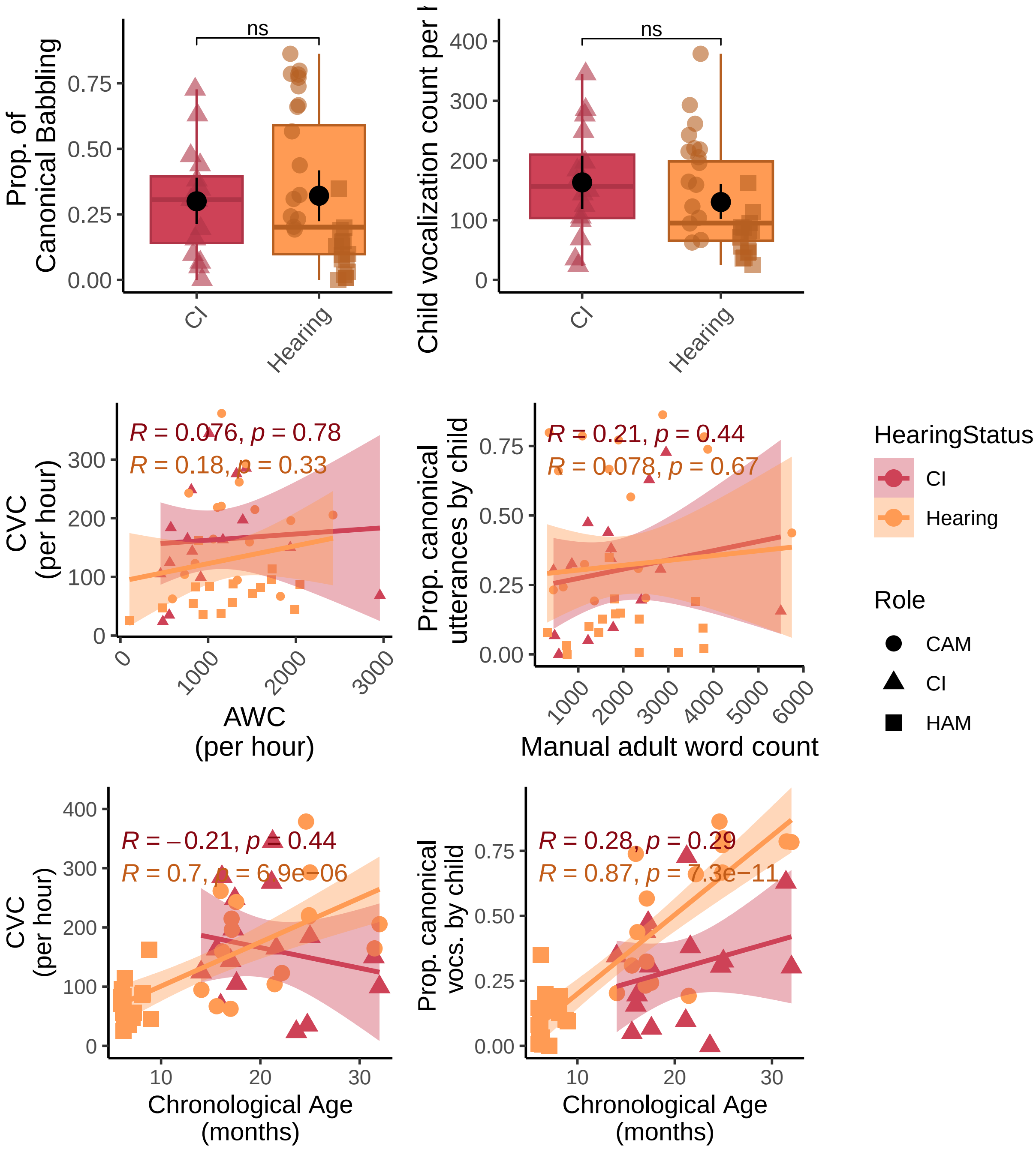

Figure 6. Input-Outcome relationships.

Lastly, we measured whether characteristics of children's language *input* predicted their language *output*. We focused just on the relationship between parent input quantity and child input quantity / maturity, instead of testing each of the input variables above, but interested readers can access our data at OSF [link] and test other possible links. For this analysis, we created models predicting children's language productions, with main effects of Age, Hearing Status, and input variable, and an interaction between that input variable and hearing status.

We started by looking at Child Vocalization Count ~ Age + $\text{AdultWordCount}_{\text{Manual}}$ + HearingStatus + $\text{AdultWordCount}_{\text{LENA}}$:HearingStatus. This model significantly predicted ~18% of the variance in child vocalization count ($R^2{}_{\text{adjusted}} = 0.18$, $p = .012$). As expected, older children produced more vocalization counts (*Beta* = 5.42, $p = .002$), but we did not find significant effects of group (*Beta* = -5.55, $p = .924$), amount of adult words in the input (by the LENA automated count) (*Beta* = 0.01, $p = .669$), or the interaction between adult word count and group. (*Beta* = 0.01, $p = .901$).

Next, we analyzed whether the proportion of canonical utterances in the child's speech was predicted by Count ~ Age + $\text{AdultWordCount}_{\text{Manual}}$ + HearingStatus + $\text{AdultWordCount}_{\text{Manual}}$:HearingStatus. This model significantly predicted ~59% of the variance in child vocalization count ($R^2{}_{\text{adjusted}} = 0.59$, $p < .001$). As expected, older children produced more canonical utterances (*Beta* = 0.03, $p < .001$). We also observed that hearing children produced a higher proportion of canonical utterances (*Beta* = 0.286, $p = .005$), and children who were exposed to more words produced a higher proportion of canonical utterances (*Beta* = 0.0001, $p = .047$). We did not find an interaction between adult word count and group (*Beta* = 0.000, $p = .271$).

Lastly, we analyzed whether the proportion of *lexical* utterances in the child's speech was predicted by Age + $\text{AdultWordCount}_{\text{Manual}}$ + HearingStatus + $\text{AdultWordCount}_{\text{Manual}}$:HearingStatus. This model significantly predicted ~67% of the variance in child vocalization count ($R^2{}_{\text{adjusted}} = 0.67$, $p < .001$). As expected, older children produced more lexical utterances (*Beta* = 0.026, $p < .001$). We also observed that hearing children produced a higher proportion of lexical utterances (*Beta* = 0.36, $p < .001$), and children who were exposed to more words produced a higher proportion of lexical utterances (*Beta* = 0.0001, $p = .014$). There was no interaction of adult word count and group (*Beta* = -0.0001, $p = .073$).

Bratt, J., Harmon, J., & Learning, B. F. &. W. P. G. L. D. M. (2022). *Morphemepiece: Morpheme tokenization.* Retrieved from https://CRAN.R-project.org/package=morphemepiece

Brugman, H., & Russel, A. (2009). Annotating multimedia / multi-modal resources with ELAN. *Proceedings of the Fourth International Conference on Language Resources and Evaluation.*

Ganek, H., & Eriks-Brophy, A. (2016, November). *The language ENvironment analysis (LENA) system: A literature review.* 2432. Umeå, Sweden: LiU Electronic Press. Retrieved from https://aclanthology.org/W16-6504

Gilkerson, J., & Richards, J. (2008). *The LENA natural language study.* Boulder, CO. Retrieved from https://www.lena.org/wp-content/uploads/2016/07/LTR-02-2_Natural_Language_Study.pdf

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (n.d.). *An open-source voice type classifier for child-centered daylong recordings.* https://doi.org/10.48550/arXiv.2005.12656

MacWhinney, B. (2019). *CHAT Manual.* https://doi.org/10.21415/3MHN-0Z89

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and Diversity: Simulating Early Word Learning Environments. *Cognitive Science*, *42 Suppl 2*(Suppl 2), 375–412. https://doi.org/10.1111/cogs.12592

Sloetjes, H., & Wittenburg, P. (2008, May). *Lrec 2008* (N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias, Eds.). Marrakech, Morocco: European Language Resources Association (ELRA). Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf

Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and MetaCorpus for studying infants' real world language experience. *Collabra: Psychology*, *7*(1), 23445. https://doi.org/10.1525/collabra.23445

Wang, Y., Cooke, M., Reed, J., Dilley, L., & Houston, D. (2022). Home auditory environments of children with cochlear implants and children with normal hearing. *Ear and Hearing*, *43*(2), 592. https://doi.org/10.1097/AUD.0000000000001124

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA language environment analysis system in young children's natural home environment* (pp. 1–16). Boulder, CO. Retrieved from https://www.lena.org/wp-content/uploads/2016/07/LTR-05-2_Reliability.pdf

Zimmerman, F. J., Gilkerson, J., Richards, J., Christakis, D., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by Listening: The Importance of Adult-Child Conversations to Language Development. *Pediatrics*, *124*(1), 342–349. https://doi.org/10.1542/peds.2008-2267