THE EVERYDAY SPEECH ENVIRONMENTS OF PRESCHOOLERS WITH AND WITHOUT COCHLEAR IMPLANTS

Margaret Cychosz[1], Rochelle S. Newman[2], Benjamin Munson[3], Rachel R. Romeo[2], Jessica Kosie[4], and Jan R. Edwards[2]

[1]University of California, Los Angeles, [2]University of Maryland-College Park, [3]University of Minnesota, [4]Princeton University
mcychosz@umd.edu

ABSTRACT

Children learn speech sounds from the language used around them. Yet children with severe to profound deafness, who receive cochlear implants, have a different learning experience. They must learn speech on a different timescale, owing to the *absence* of auditory input pre-implantation, and the *degraded* speech signal transmitted by the cochlear implant post-implantation. These differences likely shape the everyday speech and language that children with cochlear implants are exposed to. This paper analyzes daylong audio recordings of 3- to 5-year-old children with cochlear implants (16 hrs/child), and their typical hearing peers, as they go about their daily lives to characterize differences in the *quantity*, *consistency*, and experience-related *growth* in their speech-language environments.

Keywords: speech, acoustics, first language acquisition, cochlear implants, language input

1. INTRODUCTION

Children learn the sounds and structure of their native language(s) from the input that they receive from caregivers around them. Even within North American children of the same age, exposed to the same language, there is large variability in the type and quantity of speech-language exposure such as the number of word types [1], ratio of male to female speech input [2], and phonetic realization of consonants and vowels [3, 4]. Crucially, these individual differences in exposure relate to some aspects of children's speech-language development. For example, 6- to 18-month-olds who engaged in more contingent vocal interactions with caregivers vocalized more and knew more words at 18 months [5]. Other work has demonstrated similar links between the phonetics of speech input and children's outcomes: expanded vowel spaces in speech directed to 18-month-olds predicts receptive and expressive vocabulary size at 24 months [6].

There are reasons to believe that the speech-language input directed to children may vary by hearing status, particularly for children with severe to profound hearing loss who receive cochlear implants (CIs). A CI is an auditory prosthesis consisting of an audio processor worn external to the ear and an electrode array inserted into the cochlea that directly stimulates the auditory nerve, partially restoring the sensation of hearing. There are two components of cochlear implantation that may systematically alter child implantees' speech-language environments. First, children often do not receive one or both of their CIs until their first or second birthdays, resulting in a period of AUDITORY ABSENCE pre-implantation. Second, CI electrodes stimulate the cochlea at discrete points, discretizing the speech envelope. This, in addition to issues inherent to the hardware such as electrode interaction and interaural mismatch, results in a DEGRADED SPEECH SIGNAL post-implantation.

Thus, both auditory absence pre-implantation and signal degradation post-implantation characterize the listening experience of children with CIs; these perceptual changes may lead to other changes in the input, shaping these children's everyday speech and language environments. For example, among children with aided hearing loss (hearing aids), those with lower better-ear pure tone averages tend to receive more language input from adult caregivers [7]. And a series of works now demonstrate how these individual differences in linguistic input impact speech-language outcomes among children with CIs in particular [8, 9, 10, 11]. Consequently, the goal of the current study is to evaluate how the listening experiences of children with CIs shapes their everyday the speech and language environments.

2. METHODS

2.1. Participants

Eighteen children with cochlear implants participated in this study. The children were matched by parent-reported gender, socioeconomic status (instantiated as number of years of maternal education), and age to two groups of children with typical hearing (TH): (1) by *chronological* age, to match for cognitive and articulatory development (N=18), and (2) by *hearing* age, to match for auditory experience (N=16 as 2 children with CIs had <1 year of hearing experience). All children were monolingual English speakers and age matching was made within 3 months whenever possible.

The children with CIs had profound deafness in both ears (N=14 bilateral CIs, N=2 unilateral, and N=2 bimodal CI+hearing aid). All children with TH had parent-reported typical speech-language development at the time of participation. See Table 1 for demographic information by hearing group. The average maternal education level was a college degree.

Table 1: Demographic and audiological information. Mean (SD), range. *Includes the 2 children with hearing ages < 12 mos.

	Chrono. age matches	Cochlear implant	Hearing age matches
Chrono. Age (mos)	46.28 (10.8), 32-66	47.72 (9.84), 31-65	35 (12.71), 17-52
Gender (F,M)	9,9	9,9	9,7
Hearing Age (mos)	NA	31.28 (14.3), 8-54*	NA
Implant Age (mos)	NA	16.44 (9.7), 7-45	NA

2.2. Recording procedure

Each child completed one daylong recording where they wore a small, lightweight Language ENvironment Analysis (LENA) recording device (2”x3”; 2 oz.) in a specialized vest for an entire day. Recordings were completed on a typical, non-school day in the child’s life. Families were instructed to turn the device on in the morning when the child awoke and continue recording for the duration of the device battery (16 hrs.). One family instead completed a 12.83 hour recording.

2.3. Audio processing

Measures of the children’s home speech-language environments were semi-automatically derived from each child’s recording using LENA’s diarization algorithm which assigns speaker tags, timestamps, and intensity/sound pressure levels to audio clips [12]. To minimize inter-recorder differences, LENA hardware captures intensity in dBC SPL (flat frequency response across the speech range). The algorithm reports intensity measures in dBFS (maximum of 0) so we offset each value by +97dB to facilitate interpretation. All speech clips tagged as “Target child near”, “Male adult near”, and “Female adult near” were extracted. Word token count estimates from the adult clips were likewise extracted. We filtered out target child clips that contained cries, adult male and female clips that contained any non-speech, and adult male and female clips > 10s (574 clips, 0.07%). Finally, algorithmic estimates of “conversational turns”, defined as target child and adult utterances spoken within 5 seconds, were extracted (see Figure 1). This workflow allowed us to assess the quantity, consistency, and experience-related growth of the children’s speech-language environments.

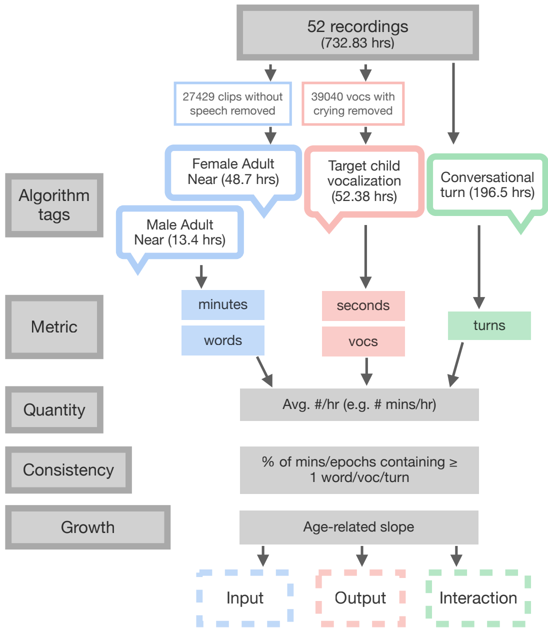

Figure 1: Daylong audio recording processing steps.

There has been substantial work evaluating the LENA system’s algorithmic performance [13]. Crucially, our analyses relied on diarization and tags that have relatively high recall and precision for the language and age group studied (e.g., “Female adult near” > 60% for English-learning infants and preschoolers [14]) and not those categories, such as electronic speech, that have poorer reliability.

Nevertheless, as algorithmic performance is not exact for any of the analyzed categories, we interpret our results by comparing *across* hearing groups (as there should be no reason why algorithmic performance would be better or worse by hearing group), and stress that reports of exact amounts of e.g., words or vocalizations per hour, should be interpreted with caution.

3. RESULTS

We divide the results section into three components of each child's daily speech-language experience—caregiver input, target child output (production), and parent-child interaction—and evaluate the impact of hearing group upon each outcome. See tables 2 and 3 for summary statistics of the measures.

Table 2: Measures of the naturalistic speech environment, by hearing group. Mean(SD), range.

	Chrono. age matches	CI	Hearing age matches
Recording duration (hrs)	15.82(0.75), 12.83-16	16(0), 16-16	16(0), 16-16
Input			
Adult speech intensity	68.21(5.97), 48.98-83.7	68.87(5.81), 45.46-84.03	68.92(6.15), 48.08-87.94
Adult speech/hr (words)	1081.49(481.29), 285.63-2250.39	1217.23(508.87), 411.36-2127.7	1105.54(433.01), 170.88-1630.86
Adult speech/hr (s)	258.36(118.69), 72.79-533.67	288.52(121.52), 95.31-499.53	264.55(102.57), 43.67-386.22
Adult word consistency	0.52(0.13), 0.31-0.78	0.58(0.11), 0.25-0.7	0.51(0.12), 0.19-0.66
Output			
Voc. intensity	76.78(4.39), 47.51-84.77	76.95(4.35), 43.16-85.79	77.15(4.95), 44.79-90.31
Child voc. quantity	308.03(142.81), 90.12-575.81	271.75(69.23), 48.75-381.62	254.5(108.83), 42.5-424
Voc. duration (ms)	1004.46(662.3), 80-10940	937.93(569.76), 80-13270	966.59(627.6), 80-19730
Child voc. consistency	0.55(0.15), 0.34-0.84	0.58(0.13), 0.17-0.72	0.49(0.14), 0.22-0.69
Interaction			
Convo. turn quantity	61.71(32.78), 20.69-150.94	68.17(26.47), 8.5-116.75	65.13(25.47), 11.12-92.62
Convo. turn consistency	0.58(0.14), 0.38-0.84	0.64(0.13), 0.22-0.77	0.56(0.12), 0.36-0.74

Data were analyzed in the RStudio computing environment (R version 4.2.1; [15]). Visualizations were made using `ggplot2` [16] and modeling was conducted using `lme4` [17] and `lmerTest` packages [18]. All model fitting began with a baseline, random-effects only model. Model fit improvements were evaluated by comparing model log-likelihood values and AIC estimates. The predictor Hearing Group (3-levels: CI, chronological age matches, hearing age matches) was contrast-coded with 'CI' as the reference level so model coefficients for the chronological and hearing age match groups refer to deviance from the CI group unless noted otherwise.

3.1. Input

Children's speech-language input was quantified as the average number of minutes/hour containing speech from an adult female or male near the child, as well as the average number of words spoken by an adult near the child/hour. Throughout the results we will normalize by hour to account for time-of-day differences across recordings, as well as different recording lengths.

For repeated measures, such as speech intensity, we fit linear mixed effects models with random intercepts by child and a fixed effect of hearing group. Models of hourly measures (words, minutes) additionally included random intercepts by hour of recording. There were no reliable differences by hearing group for speech input intensity, or measures of input quantity (hourly words, hourly minutes of speech; log-likelihood tests all $p > .05$); thus, speech input was produced at a similar intensity and quantity (minutes and words) regardless of the child's hearing experience.

We next evaluated the consistency of speech input by hearing group, or the percentage of minutes in each recording containing ≥ 1 word from an adult. There were no differences in speech input consistency by hearing group ($p > .05$), although speech input did become more consistent with age (coded continuously, in months) across the sample, independent of hearing status (model fit: β=0.004, t=3.16, p=.003) meaning that as children aged, speech was more continuously present throughout their day.

Finally, we evaluated differences by hearing group in age-related growth of speech input. For this analysis, we modeled the effect of age (in mos) upon hourly adult word token count in the children's environments and compared the values across all three hearing groups (CI, chronological age matches, hearing age matches). For the children with CIs, we modeled both their growth by *chronological* age as well as *hearing* age (time since implantation): hourly word counts only increased in the TH groups (Table 3).

3.2. Output

Each child's speech production was quantified as the average number of vocalizations/hour from

Table 3: Relationship between age (mos) and measures of the naturalistic speech environment, by hearing group. β=model coefficient, p-value from model parameter (***$p \leq .001$, **$p \leq .01$, *$p \leq .05$), +$p \leq .1$, r=Pearson correlation coefficient. No p-value annotation indicates $p \geq .1$.

	Chrono. age matches	CI chrono. age	CI hearing age	Hearing age matches
Adult words	β=20.7+ r=0.46	β=3.81 r=0.07	β=0.77 r=0.02	β=16.49+ r=0.48
Child voc. quantity	β=3.98 r=0.34	β=2.71 r=0.33	β=1.06 r=0.17	β=8.03*** r=0.89
Child voc. duration (ms)	β=.59 r=0.02	β=3.16* r=0.05	β=1.84 r=0.03	β=6.59** r=0.11
Convo. turn	β=0.83 r=0.29	β=0.01 r=0	β=-0.33 r=-0.16	β=2.18*** r=0.93

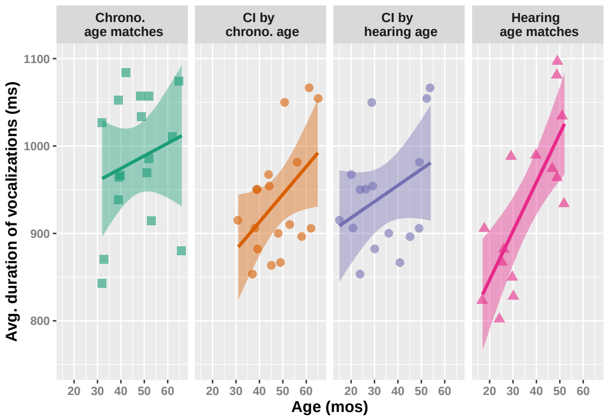

Figure 2: Age-related growth in child vocalization duration, by hearing group.

the target child and the average duration of each vocalization. There was no effect of hearing status on the number of vocalizations/hour or speech output intensity (both tests p >.05); so, hearing status did not dictate the amount or intensity of the children's speech. However, there was an effect of hearing status in the model predicting vocalization duration (χ^2=6.95, df=2, p=.03): the chronological age matches produced significantly longer vocalizations than the children with CIs (β=59.25) and hearing age matches (β=78.71).

We additionally measured consistency of speech output as the percentage of minutes in each recording containing at least one vocalization from the target child; there was no effect of hearing experience upon output consistency. And finally, we measured the age-related growth in vocalization quantity and duration: there was a significant, positive effect of age on vocalization duration among the children with CIs by chronological age and for the hearing age matches (Figure 2).

3.3. Interaction

Finally, we compared caregiver-target child interactions. We instantiated quantity of interaction as the average number of back-and-forth conversational turns per hour and the consistency of interaction as the percentage of 5-minute epochs containing at least 1 conversational turn. There was no effect of hearing group upon the quantity or consistency of turns (both p > .05). The age-related growth analysis showed increases in conversational turns for the hearing age matches.

4. DISCUSSION

This paper analyzed naturalistic daylong audio recordings from preschoolers to evaluate how the daily speech-language environment varied by hearing experience. Broadly, results show minimal differences by hearing status in caregiver-child interaction and speech input, even when we instantiated these measures in different manners (e.g., quantity, consistency, growth).

Differences by hearing group instead emerged in the children's own vocal productions: vocalization duration increases by chronological and hearing age among the children with CIs. Going forward it will be important to evaluate how different parameters of vocal production emerge in the speech of children with CIs (and how this compares to children with TH); for example, the rate of canonical babble, fluidity of consonant-vowel transitions, and number of syllables per utterance are all important indices of speech-language development. And these early comparative results suggest that they may exhibit substantial differences by hearing experience.

5. CONCLUSION

Children with CIs have a different early speech-language experience owing to the absence of auditory information pre-implantation and the degraded auditory signal conveyed by the implant post-implantation. Nevertheless, using hundreds of hours of audio capturing children's naturalistic speech-language environments, we were able to show that children with CIs and TH have fairly similar early speech-language experiences. The dynamics of the children's own vocal production, however, did vary by auditory experience, suggesting that future work should focus on the children's own speech dynamics.

6. ACKNOWLEDGEMENTS

The authors wish to thank the children and their families who participated in this research. Additional thanks to Julianna Gross for assistance with data collection and processing. All audio processing scripts and markdowns to replicate these results are included in the project's Github repository: https://github.com/megseekosh/everyday_CI.

7. REFERENCES

[1] M. L. Rowe, "A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development: Child-directed speech and vocabulary," *Child Development*, vol. 83, no. 5, pp. 1762–1774, Sep. 2012.

[2] E. Bergelson, M. Casillas, M. Soderstrom, A. Seidl, A. S. Warlaumont, and A. Amatuni, "What do North American babies hear?: A large-scale cross-corpus analysis," *Developmental Science*, vol. 22, no. 1, p. e12724, 2019.

[3] L. C. Dilley, A. L. Millett, J. D. Mcauley, and T. R. Bergeson, "Phonetic variation in consonants in infant-directed and adult-directed speech: the case of regressive place assimilation in word-final alveolar stops," *Journal of Child Language*, vol. 41, no. 1, pp. 155–175, Jan. 2014.

[4] M. Kalashnikova and D. Burnham, "Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions," *Journal of Child Language*, vol. 45, no. 5, pp. 1035–1053, Sep. 2018.

[5] N. Ferjan Ramirez, S. R. Lytle, and P. K. Kuhl, "Parent coaching increases conversational turns and advances infant language development," *Proceedings of the National Academy of Sciences*, vol. 117, no. 7, pp. 3484–3491, Feb. 2020.

[6] K. M. Hartman, N. B. Ratner, and R. S. Newman, "Infant-directed speech (IDS) vowel clarity and child language outcomes," *Journal of Child Language*, vol. 44, no. 5, pp. 1140–1162, Sep. 2017.

[7] M. VanDam, S. E. Ambrose, and M. P. Moeller, "Quantity of parental language in the home environments of hard-of-hearing 2-year-olds," *Journal of Deaf Studies and Deaf Education*, vol. 17, no. 4, pp. 402–420, Oct. 2012.

[8] L. C. Dilley, E. A. Wieland, M. Lehet, M. K. Arjmandi, D. Houston, and T. R. Bergeson, "Quality and quantity of infant-directed speech by maternal caregivers predicts later speech-language outcomes in children with cochlear implants," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. 1822–1822, Mar. 2018.

[9] L. C. Dilley, M. Lehet, E. A. Wieland, M. K. Arjmandi, M. Kondaurova, Y. Wang, J. Reed, M. Svirsky, D. Houston, and T. R. Bergeson, "Individual differences in mothers' spontaneous infant-directed speech predict language attainment in children with cochlear implants," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 7, pp. 2453–2467, Jul. 2020.

[10] M. Arjmandi, D. Houston, Y. Wang, and L. C. Dilley, "Estimating the reduced benefit of infant-directed speech in cochlear implant-related speech processing," *Neuroscience Research*, p. S0168010221000213, Jan. 2021.

[11] J. L. DesJardin and L. S. Eisenberg, "Maternal contributions: Supporting language development in young children with cochlear implants," *Ear and Hearing*, vol. 28, no. 4, pp. 456–469, Aug. 2007.

[12] D. Xu, U. Yapanel, and S. Gray, "Reliability of the LENA Language Environment Analysis System in young children's natural home environment," LENA Research Foundation, Boulder, CO, Technical Report ITR-05-2, 2009.

[13] M. Lehet, M. K. Arjmandi, L. C. Dilley, and D. Houston, "Circumspection in using automated measures: Talker gender and addressee affect error rates for adult speech detection in the Language ENvironment Analysis (LENA) system." *Behavior Research Methods*, vol. 53, no. 1, pp. 113–138, 2020.

[14] A. Cristia, F. Bulgarelli, and E. Bergelson, "Accuracy of the Language Environment Analysis System segmentation and metrics: A systematic review," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 4, pp. 1093–1105, 2020.

[15] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020.

[16] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag New York, 2016.

[17] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[18] A. Kuznetsova, P. Brockhoff, and R. Christensen, "lmerTest Package: Tests in linear mixed-effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.