Talk, You're On Camera! Or, Comparing Naturalistic Audio and Video Recordings of Infants

Elika Bergelson[1], Andrei Amatuni[1], Shannon Dailey[1], Sharath Koorathota[2], & Shaelise Tor[2]

[1] Duke University

[2] University of Rochester

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Elika Bergelson, 417 Chapel Drive, Box 90086. E-mail: elika.bergelson@duke.edu

Abstract

Measurements of infants' quotidian experiences provide critical information about early development. However, the role of sampling methods in providing this information is rarely examined. Here we directly compare language input from hourlong videos and daylong audio-recordings within the same group of 44 infants, at 6 and 7 months. We find far denser noun input in video- than in audio-recordings, across 12 measures of language quantity and lexical diversity, talker variability, utterance-type, and referential transparency. Although audio-recordings captured ~10x more awake-time than videos, the noun input in them was only 2-5x greater. Most notably, per unit time, videos featured more word-types and tokens, more questions but fewer declaratives, and more talkers. In contrast, videos often lacked certain input altogether, e.g. reading, singing, and fathers' speech. While we find moderate correlations across recording-types, the most common audio-recording nouns were far more consistant across families than top video-recording nouns. Thus, even when standardized per unit time, hourlong video and daylong audio-recordings provided fairly divergent pictures of the input infants hear learn from in their daily lives. We suggest short video-recordings may inflate various language input estimates, and should be used cautiously for extrapolation about common words, talkers, situations, and contexts at larger timescales. If theories of language development are to be held accountable to 'facts on the ground' from observational data, we advocate for greater care in unpacking the ramifications of sampling methods of early language input.

Keywords: keywords

Talk, You're On Camera! Or, Comparing Naturalistic Audio and Video Recordings of Infants

Introduction

Researchers have studied development by observing infants experiencing their natural habitats for over a century (Taine, 1876) (**ADD REF**: Williams, 1936). Over the past 20-30 years, written records have been increasingly supplemented with annotated audio and video recordings, which have described the linguistic, social, and physical landscape in which infants learn. Such data –often shared through repositories like CHILDES and Databrary–in turn provide a proxy for various "input" measures in theories of psycho-social, motor, and in particular, linguistic development (MacWhinney, 2001).

Furthermore, recent technological advances have made it feasible to collect longer, denser, and higher-quality recordings of infants' day-to-day lives, which aim to provide better approximations of infants' input and early language abilities (Bergelson & Aslin, 2017, B. C. Roy, Frank, DeCamp, Miller, and Roy (2015), Oller et al. (2010), Weisleder and Fernald (2013)) (**CHECK REF**: Bergelson & Aslin, 2017b„ Roy et al., 2015, Oller et al., 2010, Weisleder & Fernald, 2013, VanDam et al, 2016, inter alia.) Such naturalistic data seeks to reveal what infants actually learn from as they make use of their biological endowments and environmental resources.

While cutting edge technologies make collecting observational data ever easier, this growing toolbox increases researchers' decision load, with serious but underexplored side-effects. For instance, researchers must decide on recording modalities (e.g. audio, video, both), where, whom, and how long to record, and whether to capture structured or free-ranging interactions, with or without experimenters present. While any path through such decision-trees may lead to equivalent results, this is rarely tested directly. Problematically, this leads to research with theoretical conclusions built on equivalency assumptions that go unmeasured.

In recent work directly comparing observational sampling methods, Tamis-LeMonda, Kuchirko, Luo, Escobar, and Bornstein (2017) (**CHECK REF**: Tamis-Lemonda et al.

(2017)) analyzed mother-infant behavior in 5-minute structured interactions, and 45 minutes of free play. Home sessions were video-recorded by an experimenter and transcribed. The results showed that relative to free play, in structured interactions infants generally experienced more language both in word-quantity (i.e. tokens) and word-variability (i.e. types) per minute. They also found that language quantity across contexts correlated, and that the peak five-minutes of the naturalistic interaction was similar to the 5-minute structured interaction. They conclude that sampling must be matched with research-question, cautioning that while brief samples may be appropriate for studying individual differences, extrapolations about overall language input from short samples must be made with care.

In contrast, work by Hart and Risley (1995) (**CHECK REF**: Hart and Risley (1995)) extrapolated extensively. Based on 30 hours of data per family (collected one hour per month for 2.5 years), these researchers estimated that by age four, children receiving public assistance (n=6) heard >30-million fewer words than professional-class children (n=13). While their results highlighting SES differences certainly merited (and received) follow-up (Fernald, Marchman, & Weisleder, 2013) (**CHECK REF**: Noble et al, Fernald et al, 2013, inter alia), they have been criticized as an extreme over-extrapolation (Dudley-Marling & Lucas, 2009) (Dudley-Marling and Lucas, 2009; Michaels, 2013).

Still other research analyzes base rates of certain linguistic phenomena, to provide in-principle proof of what young children can learn from their input (Brent & Siskind, 2001) (**CHECK REF**: Tomasello, 2000; Lidz et al, 2003; Brent and Siskind, 2001). Here, the research question dictated what was deemed appropriate sampling. Problematically, for most exploratory work, "appropriate" sampling is hard to premeditate. For instance, practically any length of adult speech, across wide-ranging recording parameters will find function words (e.g. "of") at much higher rates than content words (e.g. "fork"). But for questions concerning infants' language input, it is largely unknown how methodological choices may bias our answers.

In the present study, we explore these issues, directly comparing hour-long video-recordings and daylong audio-recordings in a single sample of 44 infants, at 6 and 7 months, as part of a larger study on early noun learning. We annotated concrete nouns (generally, objects, foods, animals, or body-parts) said to infants, or said loudly and clearly in their presence. We further annotated three properties previously linked with early language learning: (1) utterance-type, which provides syntactic and situational information (cf. **ADD REFS** Naigles & Hoff, Debaryshe, Brent & Siskind) (2) referential transparency, which clarifies whether the target of the spoken word is visually appreciable (Bergelson & Aslin, 2017) (Yurovsky et al, Trueswell et al, Bergelson & Aslin, 2017; Bergelson & Swingley, 2013), and (3) talker, which lets us quantify the range of speakers infants hear (Rost & McMurray, Bergmann et al, 2016, Cogsci).

This design sets up two overarching questions. First, do features of the noun input in one video-recorded hour predict these same quantities in an entire audio-recorded day? Second, do input quantities differ once time is standardized? If the noun input is equivalent and predictive across recording-types, then researchers can freely vary their observational data collection approach with impunity. If it is not, understanding the biases of various methods is critical to ensuring our learning theories consider the data quantity and variability avilable to learners day-to-day.

Thus, our main goal was to compare language input young infants receive across four key properties (word quantity/diversity, utterance-type, referential transparency, and talker), as measured by hourlong videos and (separate) full-day audio-recordings. This seemingly methodological question has deep implications for developmental theory: we examine how sampling and aggregation approaches may alter conclusions about the linguistic input that in turn drives early development.

Methods

Participants

Participants were recruited from an existing database of families from local hospitals, or who heard about the BabyLab from friends, family, and outreach. Forty-six participants enrolled; two dropped out in the early stages of the project leaving 44 infants in the final sample. All infants were full-term (40 $\pm$ 3 weeks), had no known vision or hearing problems, and heard >75% spoken English in the home. Participants were 95% white; 75% of mothers had a B.A. or higher. The families were enrolled in a yearlong study that included monthly audio- and video-recordings, as well as in-lab visits every other month. Here we report on the home recording data from the first two timepoints (6 and 7 months) of this study, for which participants were compensated $10; see table XX.[1]

Procedures

Participants gave consent at an initial lab visit for the larger study through a process approved by the University of Rochester IRB. Questionnaires about various aspects of the family's and infant's background conducted during lab visits, not germane to the present analysis, are reported elsewhere (Bergelson & Aslin, 2017b; Laing and Bergelson, under review). Four recordings were collected for each infant: an audio- and video-recording at six and at seven months. Each recording was on a different day. See table XX.

	Video recordings	Audio recordings	In-lab visits
6 months	M=6;4, SD=3.2 days	M=6;7, SD=3.9 days	M=6;2, SD=3.7 days
7 months	M=7;2, SD=2.3 days	M=7;5, SD=3.3 days	NA

[1] We include only these timepoints because at this stage of development no infants had begun producing words themselves (which may change the input for reasons orthogonal to those we examine here), and given the broader project aims, these timepoints alone had the entire daylong audiorecording annotated.

Audio-video release forms were given to parents and collected after the audio and video recordings for the month were complete. Parents could opt to share the data with other authorized researchers and/or to have excerpts used for academic presentation. The released audio and video files can be accessed by registered researchers on Databrary.

Video-Recordings

Researchers visited infants' homes each month to video-record a typical hour of infants life from their own perspective. To achieve this, infants were outfitted with a hat or headband affixed with two small, lightweight Looxcie cameras (22g each). One camera was oriented slightly down and the other slightly up, to capture most of the infant's visual field (verified by Bluetooth with an iPad/iPhone during setup). A standard camcorder (Panasonic HC-V100 or Sony HDR-CX240) on a tripod was set up in a location that could best capture the infant. Parents were asked to move this camera with them if they changed rooms. After set-up, experimenters left for one hour.

Audio-Recordings

Audio-recordings captured a full day (up to 16 hours) of infants' language input. Parents were given vests with a small chest-pocket, and LENAs (LENA Foundation, Boulder, CO), small audio-recorders (<60g) that fit into the vest pocket. Parents were asked to put the vest and recorder on babies from when they awoke to when they went to bed (with the exceptions of naps and baths). Parents were permitted to pause the recorder at any time but were asked to keep such pauses minimal.

Data Processing

Details of our entire data processing pipeline are on our lab wiki (https://osf.io/cxwyz/wiki/home/). Videos were processed using Sony Vegas and in-house video-editing scripts. Footage was aligned in a single, multi-camera view before manual language annotation in Datavyu. Audio recordings were initially processed by LENA proprietary software, which segments and diarizes each audio file; this output was then converted to CLAN format for further processing and manual annotation. Through in-house scripts, long periods of silence were demarcated in these CLAN files (e.g. when the audio vest was removed or during naps). The CLAN files were then used for manual language annotation.

Language Annotation

Recordings were next annotated by trained researchers. The "sparse annotation" entailed marking each concrete noun heard by the child. This includes words directed to or easily overheard by the child (e.g. words directed at a sibling next to the infant), but not distant or background language (e.g. background television). We operationalized "object words" as concrete, imageable nouns (e.g. shoe, arm). For each object word, we included the word (as said by the speaker, e.g. "teethies"), and lemmatized to it's "basic level" or dictionary form (e.g. tooth), along with three properties: utterance-type, object presence, and talker. Utterance-type classified each object word utterance as declarative, question, imperative, reading, singing, short-phrase, or unclear. Short-phrase utterances include words in isolation and short, simple noun phrases (e.g. "the red ball" or "kitty's paw"). Object-presence was a binary measure of whether the object was present and attended to. Lastly, the word's talker was recorded, including live interlocutors and electronics: mother, brother, toy, etc.

We assessed intercoder reliability on a random contiguous 10% of the annotations in each file. **add reliability analysis**

Measure	Derived Count
Quantity	Noun tokens, Noun types
Speaker	Nouns from Mother, Nouns from Father, Unique Speakers
Utterance Type	Nouns in Declarative, Imperative, Question, Short-Phrase, Reading, and Singing Utterances

Measure	Derived Count
Object Presence	Nouns said when the referent was present and attended to

Results

Analysis Plan

Based on the coding scheme above, we derived 12 count measures from each recordings' annotations for each child (n=44), recording-type (audio, video), and month (six, seven). See Table XX. We then averaged the data from month six and seven to increase the precision of our input estimates, and since we have no theoretically-motivated reason to predict input differences across this 4 week span (i.e. there are no developmental or linguistic milestones typically achieved at 6-7 months.) We also normalized the count measures by recording length; further details are below. While the initial analysis plan was to conduct multi-level models with fixed effects of recording-type and random subject-level effects, nearly all such models revealed highly non-normal residuals (by visual inspection and Shapiro Test), limiting interpretation across measures. This was true even when data were log-transformed. We instead report a simple set of nonparametric analyses below.

For all recording-type comparisons, we look at whether our measures *differed* significantly (by two-tailed, paired Wilcoxon Test), and *correlated* significantly (by Kendall Rank Correlation) across the given groups. This approach lets us compare, e.g., whether the proportion of declaratives is indistinguishable in our audio and video recordings independently of whether these values are correlated across recording-types. We applied Holm's p-value adjustment for multiple comparisons (Holm, 1979), for the set of 12 Wilcoxon tests, and the set of 12 Kendall Correlations.

Count Measures, Audio- vs. Video-recordings

Before assessing how our 12 measures of noun input scaled between hour-long video-recordings and daylong audio-recordings, we analyzed recording lengths. Modally, videos were an hour (62 min, M=60.79 min, SD=6.31, R=27.9-74.9 min), and audio-recordings were 16 hours (960 min, M=858.41 min, SD=119.41, R=635-960 min), the maximum capacity of the LENA device. While audio-recordings began when children awoke, we further estimated the onsets and offsets of daytime naps by removing the "silent" portions of the recordings (see Methods). This provides an estimated upper-limit on infants' awake (i.e. non-silent) time (Mode = 654 min., M = 603min, SD=106.8, R=385.2-951 min). This comports with established norms for 6–8-month-olds in the US (**ADD REF:** Mandel et al, 2010), which are 180 minutes of daytime sleep, and 600 minutes of nighttime sleep. Infants were always awake during video recordings (save one infant, who fell asleep before the recording-hour ended; that video was stopped at sleep onset).

```
## # A tibble: 1 x 15
##   vboost_min vboost_awakemin vboost_types vboost_tokens vboost_speakers
##        <dbl>           <dbl>        <dbl>         <dbl>           <dbl>
## 1       0.07             0.1         0.31          0.25            0.43
## # ... with 10 more variables: vboost_MOT <dbl>, vboost_FAT <dbl>,
## #   vboost_d <dbl>, vboost_q <dbl>, vboost_i <dbl>, vboost_s <dbl>,
## #   vboost_r <dbl>, vboost_n <dbl>, vboost_op <dbl>, comp <chr>
```

```
## # A tibble: 1 x 15
##   vboost_min vboost_awakemin vboost_types vboost_tokens vboost_speakers
##        <dbl>           <dbl>        <dbl>         <dbl>           <dbl>
## 1       0.01            0.02         0.13          0.15             0.2
## # ... with 10 more variables: vboost_MOT <dbl>, vboost_FAT <dbl>,
## #   vboost_d <dbl>, vboost_q <dbl>, vboost_i <dbl>, vboost_s <dbl>,
```

Table 3

v_MOT	a_FAT	v_FAT	v_i	a_s	v_s	a_r	v_r
0.09	0.02	0.52	0.02	0.02	0.11	0.16	0.34

```
## #   vboost_r <dbl>, vboost_n <dbl>, vboost_op <dbl>, comp <chr>
```

To examine how the hour-long video data "scale" to day-length data descriptively, we first divided the 12 count metrics from the videos by those from the audio-recordings for each child, to derive "video-fraction" scores. This showed that the video-recordings were ~0.07 of the length of audio-recordings, or 0.10 of the length if only "non-silent" portions of the audio-recording are included. However, rather than a concomitant 10-fold decrease in our count metrics (as would be expected if the video captured a "representative" hour of the day), the fractions are closer to across measures; see Table XX **add M/SD table**. Thus, by and large, videos had a denser concentration of nouns across our measures than did the audio recordings.

We computed video-fractions (rather than the reciprocal, i.e. audio/video) because there were more zero values for videos than audio-recordings (e.g. instances when children did not hear any nouns sung), rendering more undefined values. Indeed, over a third of children did not hear nouns in reading or from fathers on videos in either month. See Table 3.

We next normed our count values by the number of minutes in each. For example, if an infant heard 500 noun-tokens in 800 minutes of non-silent audio-recording, and 200 in 60 minutes of videos, this was normed to .62 and 3.3 noun-tokens per minute, respectively. Unlike for the video-fractions, this allows us to retain zero values, rendering more readily interpretable results across our count and proportional measures.

With the normed data, /12 of our metrics occured at significantly lower rates in audio recordings than video recordings (adjusted-p<.05). The remaining metric, number of nouns from fathers, was statistically indistinguishable across recording types (adjusted-p>.05).

Table 4

norm_meas	norm_inflation
y_op	2.90
MOT	3.00
FAT	1.10
d	1.90
q	3.10
n	2.50
s	2.30
r	2.90
i	2.60
numtypes	3.00
numtokens	2.30
numspeakers	3.90

Thus, overall, per unit time, infants heard less noun input across our metrics of quantity, talker, utterance-type and object presence in audio recordings than in videos (per unit time).

Looking next at correlations, we find that /12 metrics correlated in audio vs. video data; number of nouns per minute heard in singing did not. The size of the correlations was significant, but relatively moderate (excluding singing, M = 0.42, 0.27 - 0.57, all adjusted-p<.05). See Table 4 and Figures 3 and 4.

Noun Frequency and Prevalence

We conclude with a set of highly exploratory analyses at the word level, which aim to provide a first-pass characterization of whether audio and video recordings captured the same nouns and the same relative frequencies across words and families. The distribution of nouns in our recordings was zipfian: of the 5801unique object words (3137 lemmas) heard across months and recording types, only 2482 (960 lemmas) were heard more than once (see Figures 5 and 6).

We examined the top 100 most frequent nouns from audio- and video-recordings (n=136 due to ties, n=68 without words that occurred zero times in one recording-type). Frequency across recording-types correlated significantly (Kendall's tau: 0.39, p<.0001,) even with zero-frequency words included (Kendall's tau: 0.25, p<.0001; see Figure 7).

Finally, looking at just the top ten words by recording-type, we find that the top audio words were far more common across families than the top video words were (see Figure 8). While four of the top 10 words overlapped (baby, book, mouth, toes), the frequency of the top audio words was roughly 3-fold that of the top video words, again reflecting the higher density of video-recordings (which were 1/10 the length of audio recordings on average). Taken together, this exploratory analysis suggests that daylong audio-recordings appear to render more stable high-frequency words across families and than do video-recordings.

Discussion

Our results can be distilled to three key findings. First, infants heard relatively more nouns in the video recordings than in the audio recordings. Per minute, infants heard ~2-5x more noun input across our quantity, speaker, utterance-type, and object-presence metrics when they and their caretakers were video-recorded for an hour versus audio-recorded for a day. Second, while our metrics correlated across audio- and video-recordings, the relative rates of the most prevalent utterance types, and the quantity of unattested data varied across them. Finally, while the highest frequency words across recording types largegly overlapped and correlated (and exhibited Zipfian frequency distributions), top words from the daylong audio-recording appear to better represent the noun input across families.

Our comparison across recording-types highlighted many differences across our noun input measures, even with family and age held constant. Our quantity results also conceptually replicate and extend those of Tamis-Lemonda et al (2017). Despite numerous methodological differences (recording lengths, experimenter presence, infant age, word class analyzed), both studies find that parent talk per unit time is significantly higher in shorter recordings. While the difference they find is less extreme numerical (roughly 1.5-2x the number of types and tokens in the longer vs. shorter recording compared to our 2-3-fold difference), this general pattern appears robust across our very different sampling methods. Taken together, these results converge in suggesting that shorter recordings elicit denser caregiver talk.

We find consistently more object co-presence in video- than in audio-recordings. This may be because the video recordings truly had more object presence (i.e. infants mostly stayed in 1-2 rooms, interacting with caregivers and objects at hand). Alternatively, or additionally, it may be the case that there are more ambiguous cases of "object co-presence" in audio recordings than video recordings, which were in turn annotated as "not present" at higher rates. Given the XX rates of agreement, we find it more likely that this reflects a true difference between situations that arise during daylong-audio vs. hourlong-video recordings. Insofar as object presence is linked with early word learning (Bergelson & Aslin, 2017), a more extensive understanding of what modulates it is an important issue left to future work.

We did not anticipate that the top utterance-types would vary by recording-type. That is, while questions and declaratives made up the majority of the input for each recording-type at each month, videos had relatively more questions and fewer declaratives. This is key example of methodological choices potentially influencing language acquisition theories: base rates of interrogatives taken from videos would inflate estimates of auxiliary verbs in the early input. Indeed, previous work has noted that published studies vary in whether they find links between questions (yes/no and wh-) in the input, and children's early productions, with developmental level of the child invoked to explain differences across studies (Barnes et al, 1983; see discussion in Huttenlocher et al, 2002). Here we add the possibility that recording-type too may contribute to the base-rates of questions in the input, even with age kept constant.

Top Words

The pattern across recording-types suggests to us that parents behaved naturally during recordings, but that "natural" behavior differed by recording context. This is consistent with a point made by Suskind et al (2013) regarding an intervention: "sustaining increased talk for a 10-hr recording day is much less likely than being on best behavior during [a] 1-hr videotaped session…" While their work aimed to encourage caretakers to talk more, the point stands for our goals of observing infants' typical input. We add to their suggestion that shorter video-recording itself may elicit certain kinds of interactions, separate from deliberate intent or lack thereof on caretakers' part.

Indeed, the kinds of everyday interactions we captured in daylong audio recordings (family members rushing to get out the door or get meals on the table, sibling quibbles, etc.) tended to "feel" more natural. Families likely simply found it easier to go about their day freely with infants in a special vest than with a camera on their head, and a camcorder in the corner. Lending some support that equipment prominence matters, both "hat" and "camera" are in the top 10 words from video-recordings each month; no analogous nouns (e.g. vest, recorder) topped the frequency rankings in our audio recordings (see Figure XX).

Our interpretation of the present results is that findings based on relatively short video-recordings overestimate young infants' typical noun input, and that extrapolation based on daylong audio recordings likely better represents infants' quotidien experiences. This underscores our third main result: that the conclusions one would draw about which words are most common in young infants' language input differed in their robustness across families by recording type. That is, the top audio words were all heard by $\geq 75\%$ of the families we recorded; only one of the top 10 video words ("hat") was this common across families. This is true even though the video words had greater quantities of nouns per unit time; the top audio words only occured 2-4 times more often than the top video words (despite a 10-fold increase in awake recording time).

Limitations

Given the technical limitation that available infant-friendly video-recorders have a shorter battery life than audio-recorders at present, we cannot conclusively separate the effects of modality and length. That is, had we only audio-recorded for an hour, or recorded video all day, we may have obtained equivalent results across recording modalities. Such a comparison awaits technological progress.

A further limitation is the likely influence of self-selection into the study: many parents are unwilling to invite researchers to record their infants' interactions. Relatedly, our convenience sample does not reflect the broader demographics of the US (let alone other cultures or populations), and as such this work merits extension to other populations before conclusive generalizations about sampling methodology can be made (cf Bergelson et al, under review).

Conclusions

Understanding what infants learn from is a key part in understanding what and how they learn at all. Here we have taken first steps in understanding how two different data collection approaches may influence our conclusions about early linguistic input. We find that even naturalistic observer-free video-recordings appear to inflate language input. Without knowing how our sampling methods may be limiting us in principle, we necessarily limit our ability to adequately model infant learning. We…

Notes to self: Might want to reorder points, and make point 3 a little more general, some overarching version of the differences in conclusions Discuss zeros/unattested datapoints

References

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916–12921.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.

Dudley-Marling, C., & Lucas, K. (2009). Pathologizing the language and culture of poor children. *Language Arts*, *86*(5), 362–370.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children.* Paul H Brookes Publishing.

Oller, D. K., Niyogi, P., Gray, S., Richards, J., Gilkerson, J., Xu, D., … Warren, S. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354–13359.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668.

Taine, H. (1876). Note sur l'acquisition du langage chez les enfants et dans l'espèce humaine. *Revue Philosophique de La France et de L'Etranger*, 5–23.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

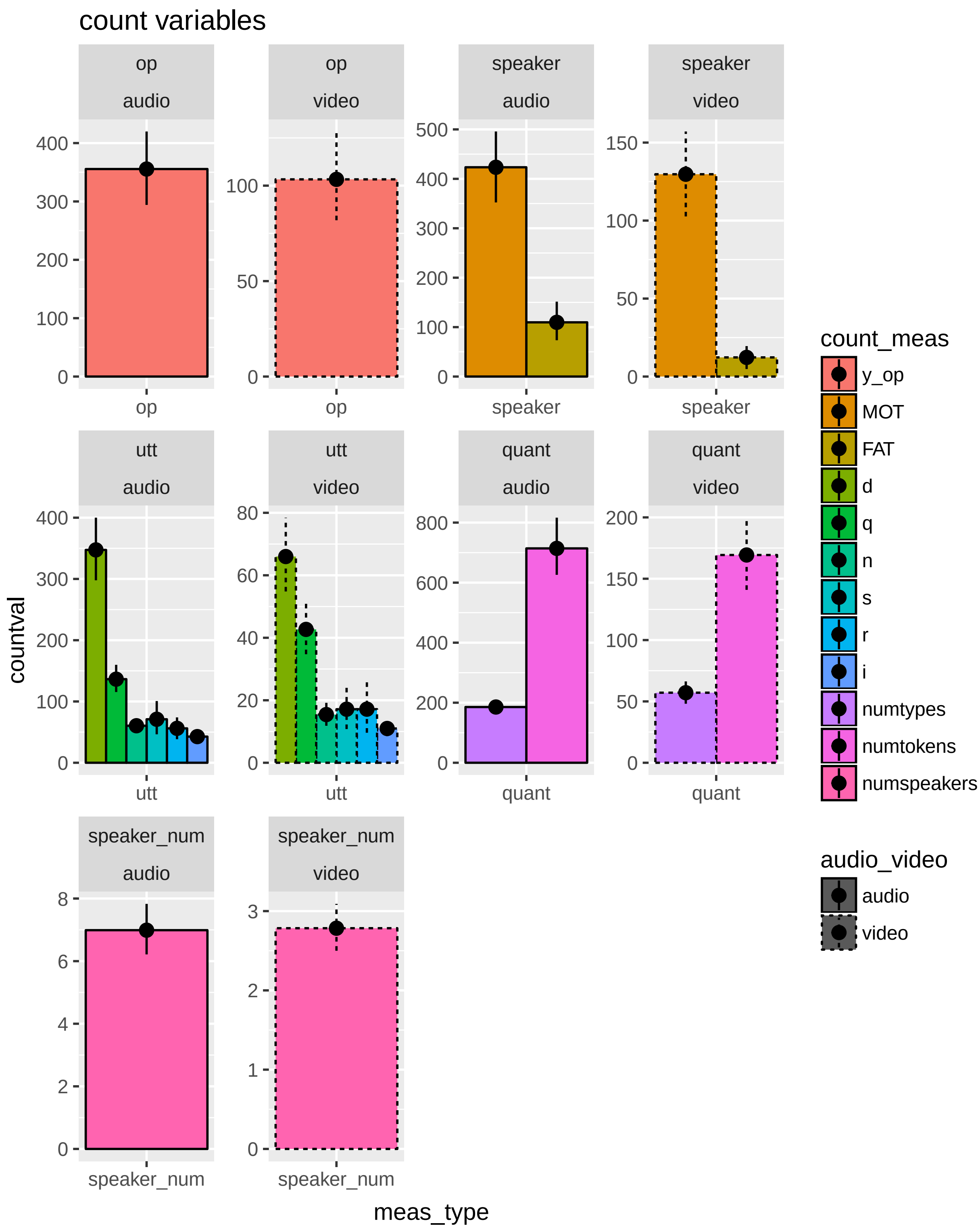

Figure 1. Counts A vs. V

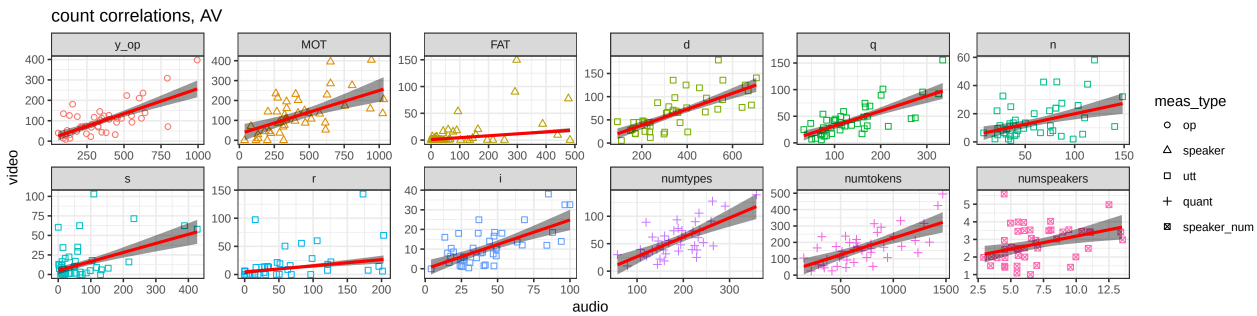

Figure 2. Count Correlations, A vs. V

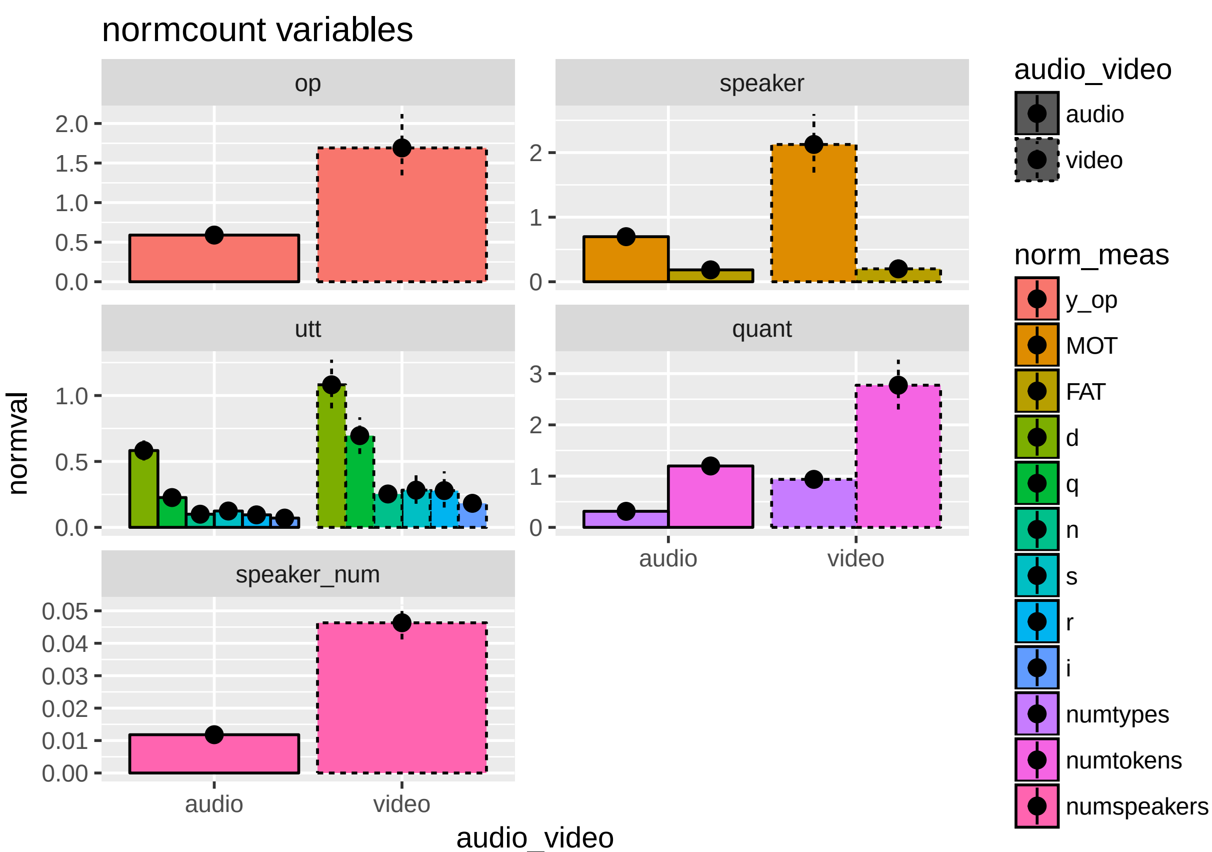

Figure 3. Normalized variable counts by month

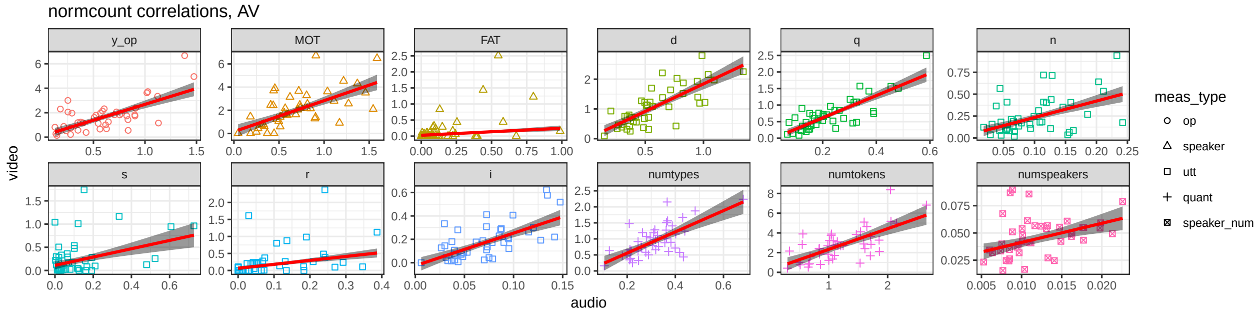

Figure 4. Normalized count correlations

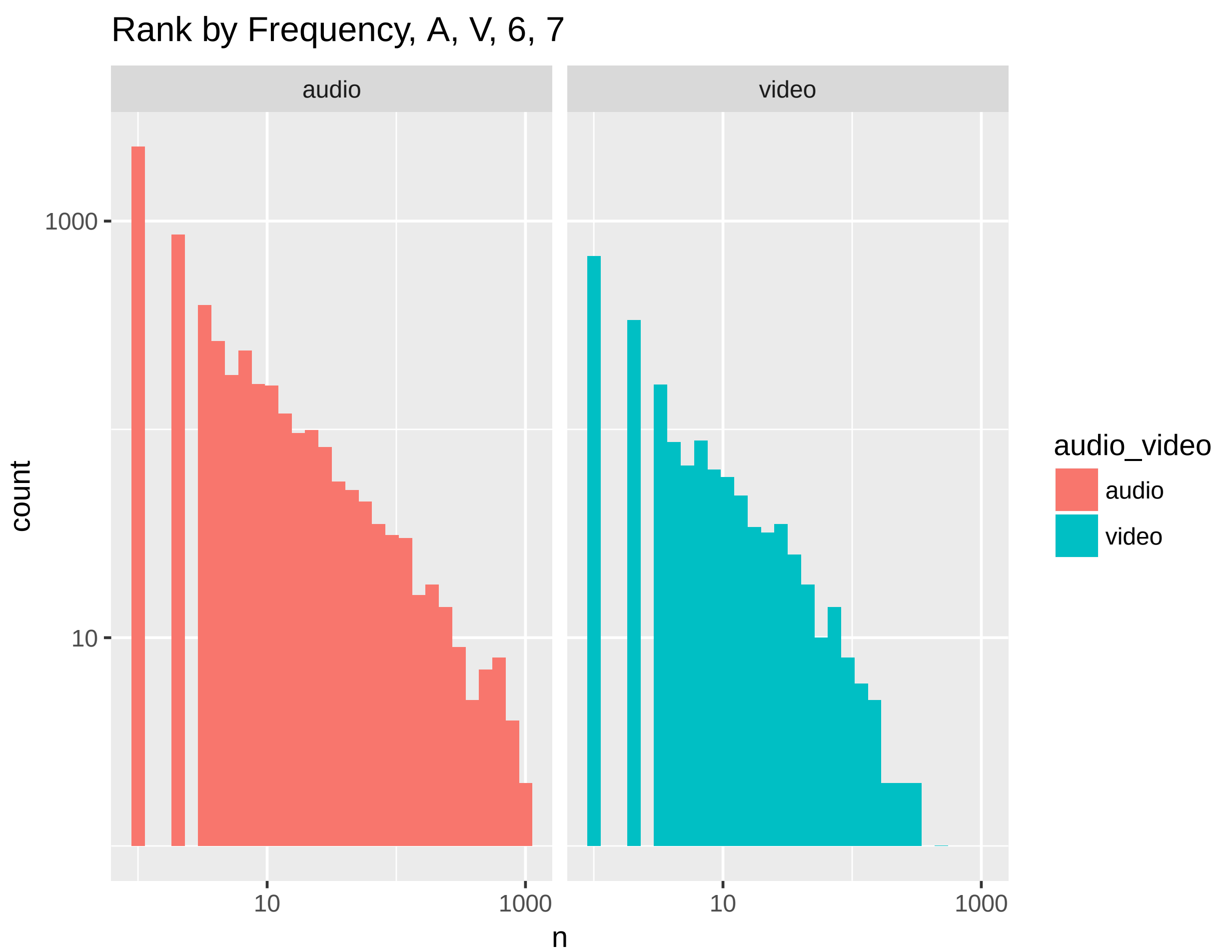

Figure 5. Zipfian word frequency distributions

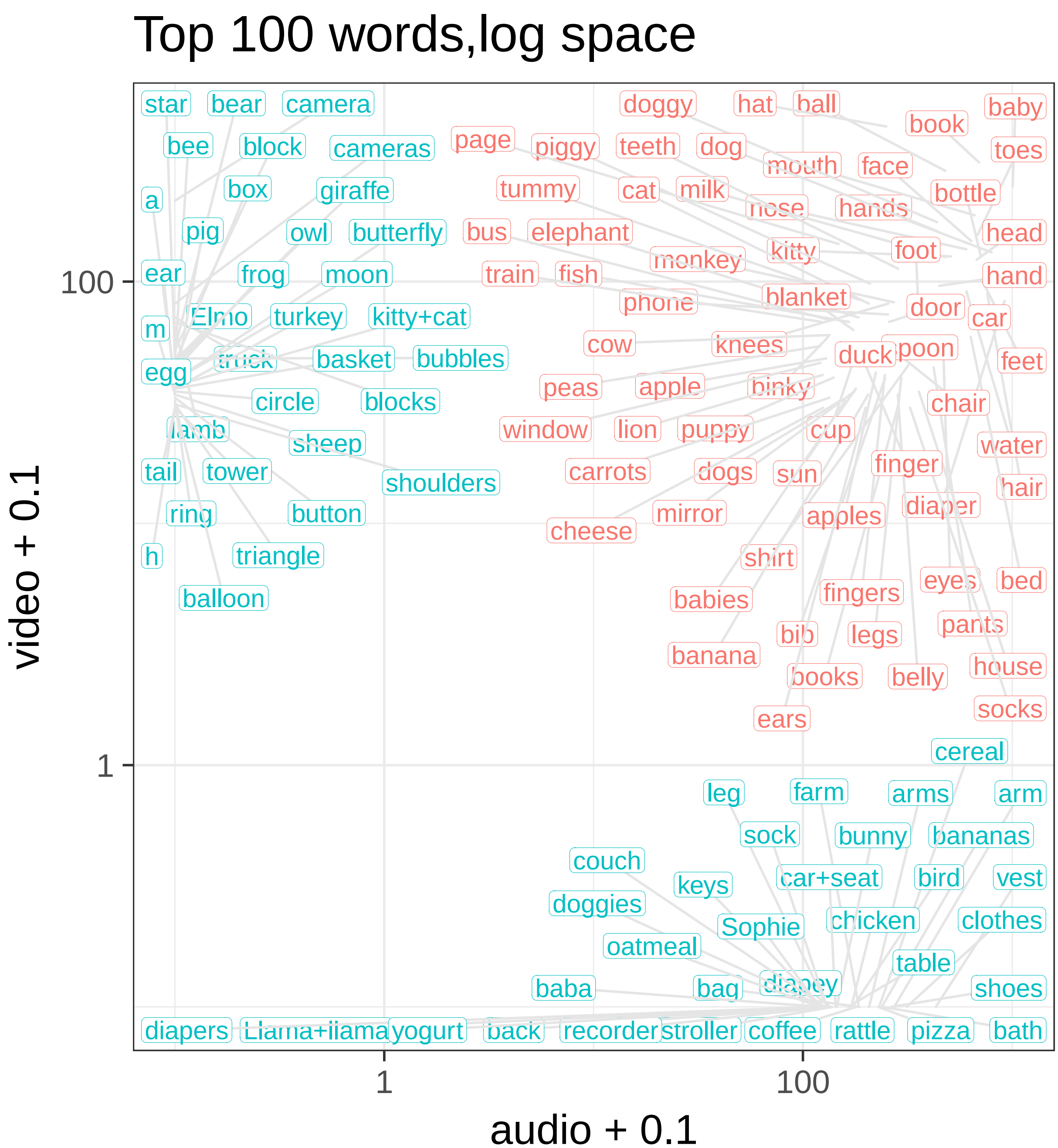

Figure 6. Top 100 words in log space

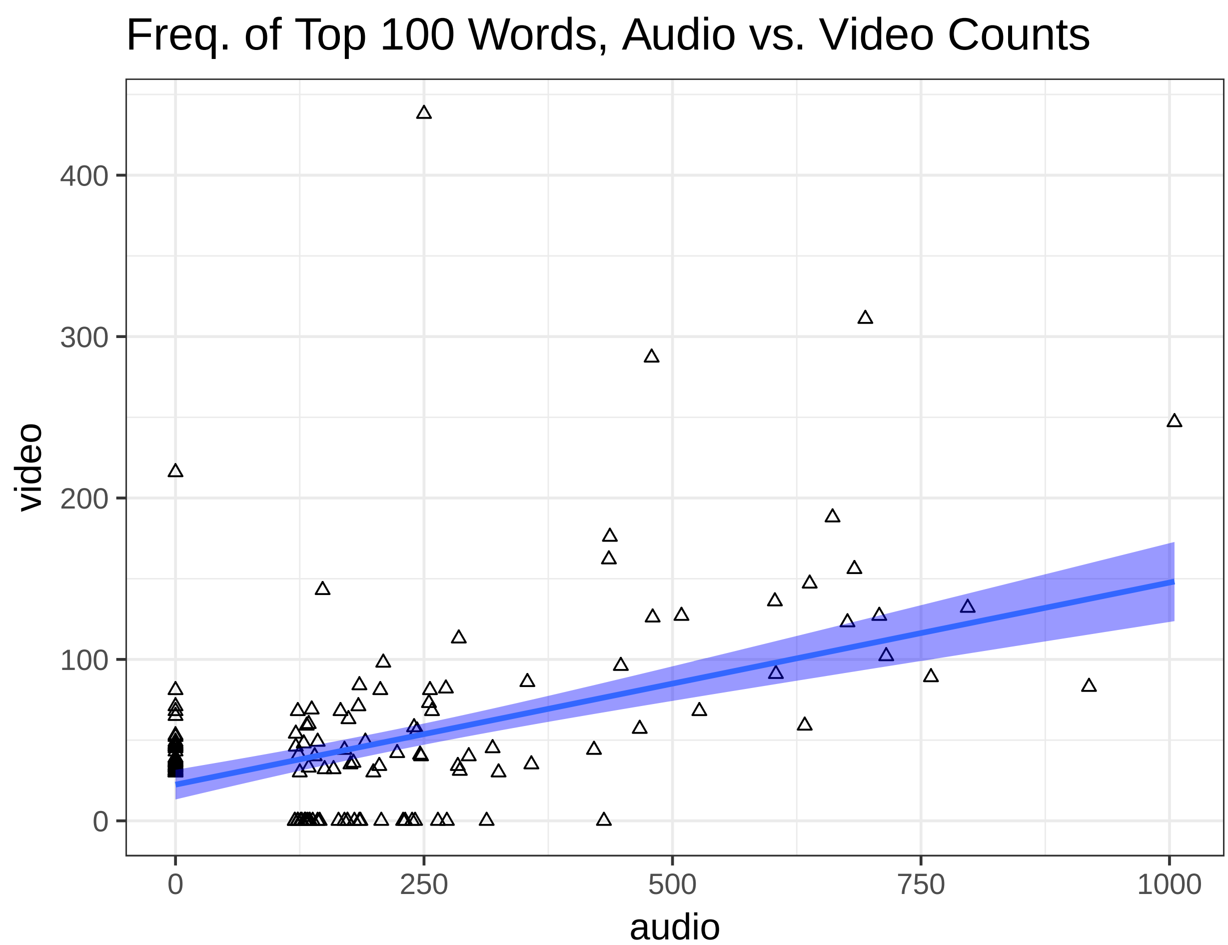

Figure 7. Top 100 words correlations by recording type

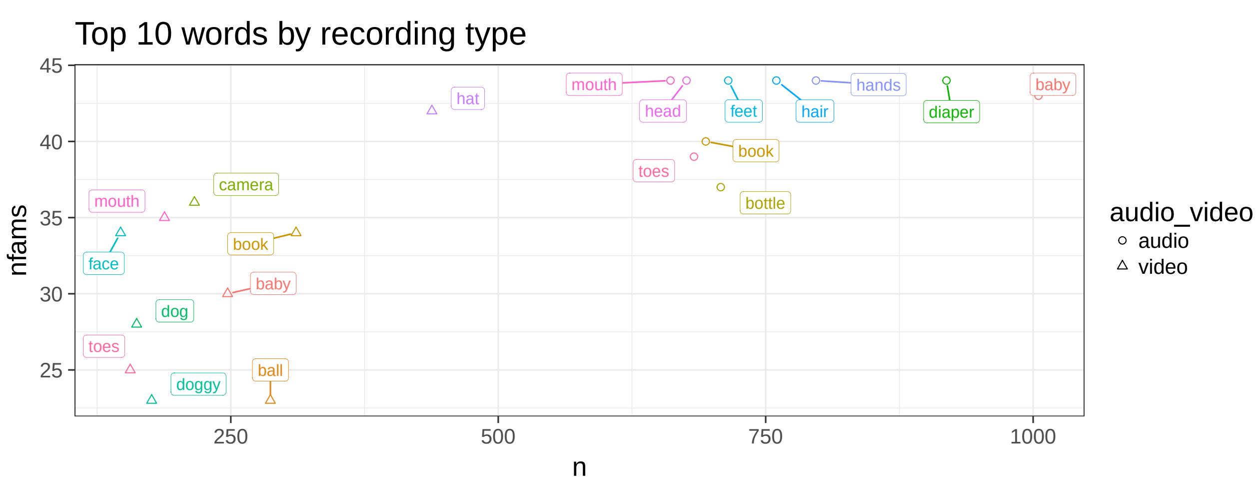

Figure 8. Top 10 words by month and recording type