**Intelligibility of single words in young children’s speech**

## Rationale

Intelligibility measures how well a listener is able to recover a speaker’s intended message from a speech signal. Research commonly examines predictors of intelligibility by looking at speaker characteristics or acoustic features. But speech perception is probabilistic, and the frequency or structure of a word can predict how it is perceived (Vitevitch & Luce, 2016). In other words, instead of asking how intelligible speakers are, it is also important to ask how intelligible individual words are. For this study, we measured the intelligibility of young children and asked how words varied in intelligibility and whether word-level characteristics (frequency, phonotactic probability, motor complexity) predicted intelligibility.

## Methods

**Participants**. Participants included 165 children (72 boys, 93 girls) between 30 and 47 months in age. Children were evenly distributed across the overall age range: 57 between 30–35 months, 51 between 36–41 months, 57 between 42–47 months. Children had no history of speech or language concerns and scored within normal limits on a speech articulation assessment.

**Task**. Speech samples were collected in a structured repetition task based on the TOCS+ (Hodge & Daniels, 2007). Prompts included 40 single words; 2 words were reserved for practice trials, so we analyzed the 38 test items. Unfamiliar listeners transcribed the children’s productions; they were played samples and instructed to type the words the child said. Every child had transcriptions by two listeners; each listener only heard productions from one child. A production was intelligible if a listener correctly transcribed the word (or a homophone).

**Analysis**. We performed a Bayesian item-response analysis (Bürkner, 2019). This mixed-effects logistic regression model estimates the expected intelligibility of an average item on an average participant (overall intercept), each participant’s ability (by-child intercept) and each item’s difficulty (by-item intercept). We included covariates to examine whether age and item-level features predicted intelligibility. These features included frequency of the word in speech, phonotactic probability, and motor complexity. Frequency counts were based on the SUBTLEX-US database (Brysbaert & New, 2009); we used frequency per million words (log10-transformed). Phonotactic probability measures were computed using the IPhOD database (Vaden, Halpin, & Hickok, 2009), using each word’s average biphone probability (log10-transformed). We computed each word’s total motor complexity score using the scale by Kuruvilla-Dugdale, Custer, Heidrick, Barohn, & Govindarajan (2018) based on Kent (1992). This systems assigns a score to each syllable part (onset, nucleus, coda) based on its articulatory motor demands, ranging from 1 (/ə, ɑ/) to 8 (cluster of 3 consonants). Table 1 reports the observed intelligibility averages and lexical characteristics for each item.

## Results

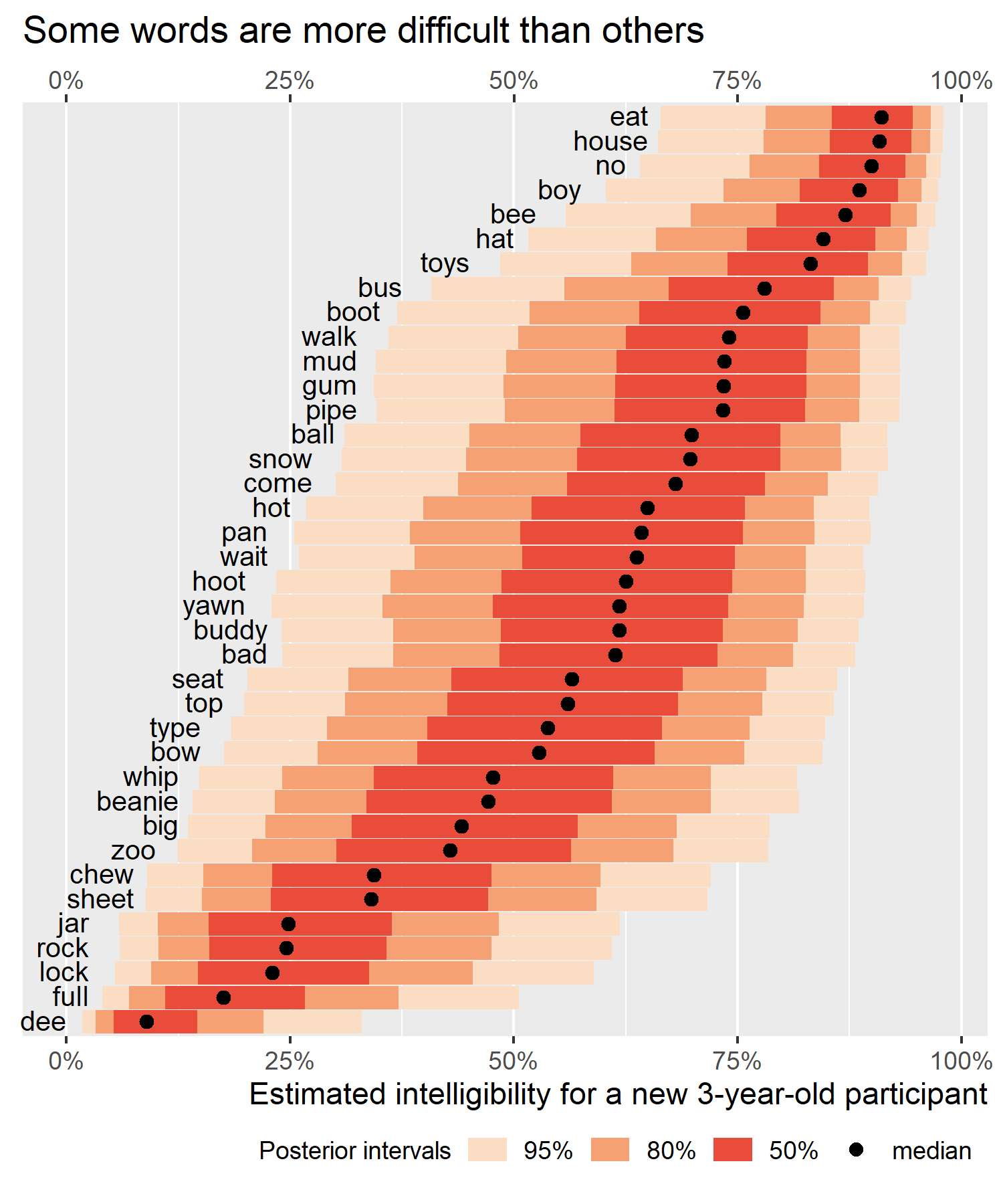
Figure 1 shows the posterior predicted intelligibility of each item for a new, unobserved 3-year-old child. There were reliable item-level differences, such that some words had a higher average intelligibility than others. Of the three word-level features, only frequency had a clear effect on average intelligibility (Figure 2). For an average item on an average 36-month-old child with average frequency (51 per million words), phonotactic probability (.0011), and complexity (9.87), the expected intelligibility was 56%, 95% interval [48, 65]. From this reference, a 10x increase in frequency predicted an increase in average intelligibility of 9 percentage points, 95% interval [0, 18]. For phonotactic probability, both positive and negative effects were equally plausible, odds ratio 95% interval [0.38, 1.66]. For a 1-point increase in motor complexity, the median odds ratio was 0.93, and the 95% interval was [0.79, 1.09]. The sign of the effect is likely negative (complexity penalty), but it is also plausible the effect is too small to be meaningful. Finally, there was a clear effect of age: A 6-month increase (from 36 to 42 months) predicted a corresponding increase in intelligibility of 12 percentage points, 95% interval [9, 15].

## Discussion

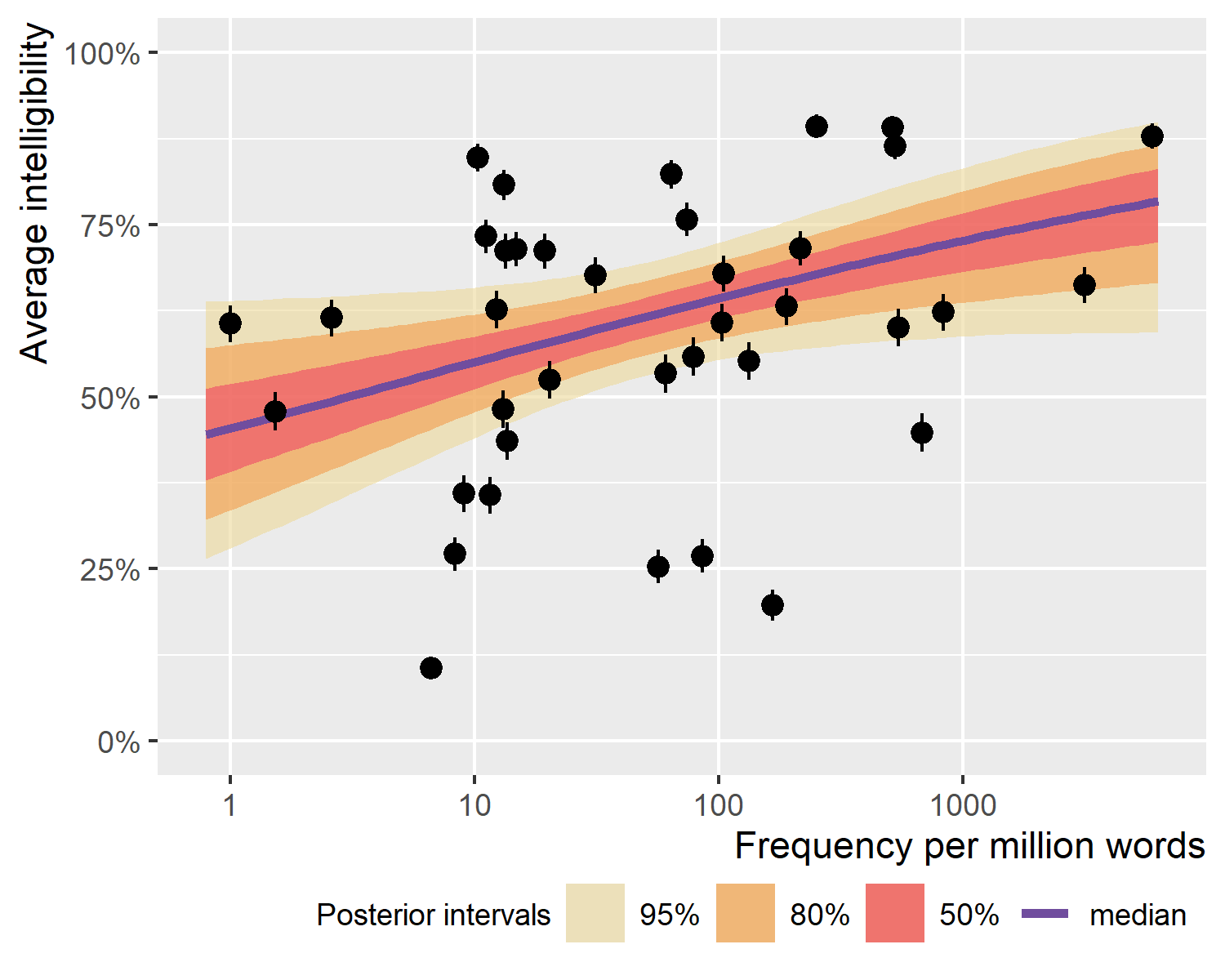
Young children’s intelligibility varied greatly at the word level. Higher frequency words tended to be more intelligible than lower frequency words, and it was unclear whether the word’s motor complexity affected intelligibility. Speakers in this study were typically developing, so further work is needed to see whether listeners rely more on word-level features when presented with disordered speech. It is plausible listeners will use more top-down knowledge from lexical statistics when hearing dysarthric children.

**Table 1.** Item-level intelligibility results and measures.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Intelligibility [95% CI] | Freq. per million [Rank] | Mean biphone probability [Rank] | Motor complexity |
| eat | 89.1% [85, 92] | 251.9 [ 8] | .00168 [13] | 7 |
| house | 89.1% [85, 92] | 514.0 [ 7] | .00035 [32] | 12 |
| no | 88.0% [84, 91] | 5971.5 [ 1] | .00089 [26] | 5 |
| boy | 86.9% [82, 90] | 529.8 [ 6] | .00013 [36] | 7 |
| bee | 84.7% [80, 88] | 10.4 [32] | .00097 [24] | 6 |
| hat | 82.5% [78, 86] | 64.2 [18] | .00155 [16] | 12 |
| toys | 80.3% [76, 85] | 13.2 [27] | .00009 [38] | 14 |
| bus | 75.9% [71, 80] | 74.2 [17] | .00685 [ 2] | 11 |
| boot | 73.7% [68, 78] | 11.1 [31] | .00093 [25] | 11 |
| gum | 71.5% [66, 76] | 13.4 [26] | .00427 [ 5] | 8 |
| mud | 71.5% [66, 76] | 14.8 [24] | .00540 [ 3] | 8 |
| pipe | 71.5% [66, 76] | 19.4 [23] | .00054 [30] | 9 |
| walk | 71.5% [66, 76] | 215.9 [ 9] | .00134 [18] | 8 |
| ball | 68.2% [63, 73] | 105.0 [13] | .00099 [23] | 12 |
| snow | 68.2% [62, 73] | 31.4 [21] | .00077 [29] | 10 |
| come | 66.0% [61, 71] | 3141.0 [ 2] | .00739 [ 1] | 8 |
| hot | 62.7% [58, 68] | 189.8 [10] | .00103 [22] | 9 |
| pan | 62.7% [57, 68] | 12.3 [29] | .00434 [ 4] | 10 |
| wait | 62.7% [57, 67] | 830.3 [ 3] | .00365 [ 6] | 12 |
| hoot | 61.6% [56, 67] | 2.6 [36] | .00085 [27] | 10 |
| bad | 60.5% [55, 65] | 545.2 [ 5] | .00131 [19] | 12 |
| buddy | 60.5% [55, 66] | 102.9 [14] | .00360 [ 7] | 11 |
| yawn | 60.5% [55, 66] | 1.0 [38] | .00053 [31] | 10 |
| seat | 56.1% [50, 61] | 78.8 [16] | .00208 [11] | 13 |
| top | 55.0% [50, 61] | 133.4 [12] | .00122 [20] | 9 |
| type | 53.9% [48, 59] | 60.6 [19] | .00083 [28] | 11 |
| bow | 52.8% [47, 58] | 20.3 [22] | .00023 [35] | 7 |
| beanie | 48.4% [42, 53] | 1.5 [37] | .00162 [15] | 11 |
| whip | 48.4% [43, 54] | 13.2 [28] | .00228 [ 9] | 10 |
| big | 45.1% [39, 50] | 682.8 [ 4] | .00167 [14] | 12 |
| zoo | 44.0% [38, 49] | 13.6 [25] | .00013 [37] | 8 |
| chew | 36.3% [31, 41] | 9.1 [33] | .00035 [33] | 8 |
| sheet | 35.2% [31, 41] | 11.6 [30] | .00110 [21] | 13 |
| jar | 27.5% [23, 32] | 8.3 [34] | .00270 [ 8] | 12 |
| rock | 26.4% [22, 32] | 86.2 [15] | .00175 [12] | 10 |
| lock | 25.3% [21, 30] | 56.6 [20] | .00147 [17] | 10 |
| full | 19.8% [16, 24] | 166.9 [11] | .00026 [34] | 13 |
| dee | 11.0% [8, 14] | 6.6 [35] | .00223 [10] | 6 |



**Figure 1.** Each item’s estimated intelligibility score for a new, unobserved participant. These posterior predictions incorporate uncertainty about the overall average intelligibility, word-level effects, variation in the population of children, and variation for specific items. Thus, there’s a 50% probability that a typically developing 30–47-month-old’s production of eat will be at least 90% intelligible and a 95% probability that full will be less than 50% intelligible.



**Figure 2.** Visualization of the frequency effect. There is a small, logarithmic effect of word frequency such that average frequency increases with 10x increases in frequency. Points and errors bars are the observed mean and standard error for each item.

## References

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>

Bürkner, P.-C. (2019). Bayesian item response modelling in R with brms and Stan. *arXiv:1905.09501 [Stat]*. Retrieved from <http://arxiv.org/abs/1905.09501>

Hodge, M. M., & Daniels, J. (2007). TOCS+ intelligibility measures [computer software]. Edmonton, Alberta: University of Alberta.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). Timonium, MD: York Press.

Kuruvilla-Dugdale, M., Custer, C., Heidrick, L., Barohn, R., & Govindarajan, R. (2018). A phonetic complexity-based approach for intelligibility and articulatory precision testing: A preliminary study on talkers with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2205–2214. <https://doi.org/10.1044/2018_JSLHR-S-17-0462>

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, version 2.0. [Data file]. <http://www.iphod.com>.

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics*, *2*(1), 75–94. <https://doi.org/10.1146/annurev-linguist-030514-124832>