Title
Corn on the What? Development of Phrase Frequency Effects in Background Noise

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Kueser, Justin B.
Zhang, Jacqueline
Thome, Emma Kate
et al.

Publication Date
2026

Corn on the *What?* Development of Phrase Frequency Effects in Background Noise

Justin B. Kueser (justin.kueser@boystown.org)[1], Jacqueline Zhang[2], Emma Kate Thome[1], Katherine R. Gordon[1], Stephanie Lowry[1], & Kaylee Cowdrey[1]

[1]Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital

[2]San Diego State University / University of California San Diego Joint Doctoral Program in Language and Communicative Disorders

Abstract

Children recognize and produce more frequent phrases (e.g., "corn on the cob") more accurately than less frequent phrases (e.g., "corn on the tray"). Here, we ask how phrase frequency affects language processing in background noise and whether this effect changes across childhood. We presented high- and low-frequency four-word phrases ("corn on the cob" vs. "corn on the tray") in quiet or in speech-shaped background noise to 40 children with typical language development aged 3-12 years. Children's repetition accuracy, response time, and phrase duration were measured. Across measures, results showed advantages for high- vs. low-frequency phrases, quiet vs. noise, and older vs. younger children. Interactions between age, frequency, and noise condition were near-zero for most measures, showing limited change across development in the effect of frequency and listening condition. Across typical development, phrase frequency aids language comprehension and production but is not preferentially used to overcome background noise.

Keywords: language production; language comprehension; frequency; background noise; development

Introduction

Adverse listening conditions with background noise are common in children's everyday lives (Erickson & Newman, 2017). How do children contend with noise during language processing? One way is by using prior language knowledge. Contemporary theories of language processing posit that language frequency information influences comprehension and production (e.g., Ambridge, 2019; Levy, 2008; Pickering & Garrod, 2013). All else being equal, material that occurs more frequently in the language input is more likely to be processed accurately (Brysbaert et al., 2018). In background noise, adults and children recognize higher-frequency words more accurately and/or rapidly than lower-frequency words (Cooke et al., 2019; Howes, 1957; Krull et al., 2010; Strauß et al., 2022; Van Engen et al., 2020).

Currently, there are two major gaps in our understanding of how language frequency affects processing in noise:

The first major gap is that work on frequency effects in background noise has focused only on single-word frequency. Studies in quiet listening conditions with no background noise show that the frequency of *multi-word phrases* also affects processing (Christiansen & Arnon, 2017). For example, in quiet, adults recognize frequent phrases like "don't have to worry" faster than infrequent phrases like "don't have to wait" (Arnon & Snider, 2010). Whether phrase frequency affects language processing in background noise is unknown. While it may support processing as it does in quiet, for reasons explained below, phrase frequency may be more difficult to use in noise.

The second major gap is that it is unknown how the effect of phrase frequency in background noise and quiet develops over childhood. This is important to learn because background noise is more detrimental for children than for adults (McCreery & Stelmachowicz, 2011; Wightman & Kistler, 2005). As phrase frequency may help scaffold higher-level language (e.g., McCauley & Christiansen, 2017), understanding how use of this knowledge develops over childhood can also help constrain theories about how language develops in real-life communicative contexts.

The goal of this study is to begin to close these gaps by measuring how phrase frequency affects language processing in background noise across childhood (i.e., 3-12 years).

Phrase Frequency Effects in Children

Children's language production is more accurate in high-frequency phrases than in infrequent or novel phrases (Arnon & Clark, 2011; Bannard & Matthews, 2008; Kueser & Leonard, 2020). For example, Bannard and Matthews (2008) examined how phrase frequency affected repetition success in 2- and 3-year-old British English-speaking children. The children heard high-/low-frequency pairs of phrases that differed only in the final word (e.g., "a cup of tea" vs. "a cup of milk"). The children's repetition attempts were scored on only the first three words of the phrases to measure the effect of the frequency of the entire phrase rather than the frequency of the final word (e.g., "tea" vs. "milk"). The first three words of high-frequency phrases were repeated more accurately and more quickly than the same words in low-frequency phrases. Kueser and Leonard (2020) extended this paradigm to American English phrases (e.g., "corn on the cob" vs. "corn on the tray"), finding that phrase frequency helped 3-to-5-year-old children with and without developmental language disorder. Arnon and Clark (2011) observed similar patterns in four-year-olds for the production of irregular plurals like "teeth" in high-frequency contexts like "brush your teeth" vs. low-frequency contexts like "so many teeth".

Development of Phrase Frequency Effects

Frequency effects in language are present from toddlerhood (Ambridge et al., 2015) through older adulthood (DeDe & Flax, 2016). Frequency information is gleaned from the language input and thus is learned through experience,

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

meaning that there are likely differences in how frequency affects language comprehension and production across development. Some work has directly tested how word frequency effects vary with age. For example, Davies et al. (2017) measured reading and writing of single words in participants aged 8-83 years. They found that the facilitatory effect of word frequency diminished across the lifespan, likely because overall facility in language processing neared asymptote for older adults, even for low-frequency material.

Current evidence for developmental changes in the effect of phrase frequency is mixed. In Bannard and Matthews (2008), the 2-year-old children had a larger frequency effect than the 3-year-old children. In contrast, in Kueser and Leonard (2020), 4-to-5-year-old children with developmental language disorder, 4-to-5-year-old children with typical language development, and a younger 2-to-3-year-old group with typical language development all showed similar effects of frequency. Similarly, Matthews and Bannard (2010) examined the effect of other statistical properties of phrases on repetition performance and found that both 2- and 3-year-olds experienced similar effects. One possibility for the differences across studies may be that the age range of the children in each study was relatively narrow, differing in only one to two years from one another. Examining phrase frequency effects across a larger age range within a single study may provide a better opportunity to detect changes.

Use of Phrase Frequency Information in Noise

In background noise, higher-frequency words are recognized more accurately and rapidly than lower-frequency words (Broadbent, 1967; Cooke et al., 2019; Howes, 1957; Morton, 1969; Pollack et al., 1960; Strauß et al., 2022; Van Engen et al., 2020). Whether this also applies to the recognition and production of multi-word phrases is unknown. On the one hand, the advantage of high frequency may be heightened in noise compared to quiet. For high-frequency phrases, as the phrase unfolds, it is likely that a child will predict the final word (e.g., "cob" from "corn on the…") before the phrase is completed (e.g., Altmann & Kamide, 1999). This pre-activation may interfere with the child's ability to recognize the last word in low-frequency phrases if the last word is hard to hear, as in background noise. In speech-shaped noise, children better recognize high semantic context sentences (e.g., "tough guys sound mean") vs. low context sentences (e.g., "quick books look bright"; Buss et al., 2018). Though this study did not control for word/sentence frequency, similar effects may be found in high-frequency phrases (that are presumably more semantically familiar). On the other hand, the effect of phrase frequency could be reduced if the child has difficulty recognizing each word of the phrase, as may occur in background noise. For example, if there is high uncertainty in identifying each word individually, recovering enough of the sentence context to identify the whole phrase may be challenging. Thus, in noise, frequency information may be more likely used for single words vs. phrases.

Additionally, we know little about how phrase frequency effects in background noise might interact with age. We know that both adults and children experience an advantage for the recognition of high-frequency words (Krull et al., 2010; Van Engen et al., 2020) but how this plays out with phrase frequency effects could be different. As children's ability to deal with adverse listening conditions like noise improves over development (Corbin et al., 2016; Walker et al., 2019), one possibility is that older children may be better able to use phrase frequency information in background noise as they will be better able to recognize each word in the phrase. Alternatively, earlier in development, listeners may be more tied to their frequency knowledge as they have not yet developed the ability to accurately listen in background noise, and so the advantage for high- vs. low-frequency phrases may be heightened for younger children.

Current Study

The goal of the current study is to examine how background noise affects the processing of high- vs. low-frequency phrases across childhood. Specifically, we ask: (1) How does phrase frequency affect processing of frequent (e.g., "corn on the cob") vs. infrequent (e.g., "corn on the tray") phrases and how does this effect change over childhood?, and (2) How does phrase frequency affect processing in background noise and how does this effect change over childhood?

To answer these questions, we adapted the phrase repetition task developed in Kueser and Leonard (2020). Repetition tasks are highly sensitive to language processing (Archibald & Joanisse, 2009), making them ideal for testing a wide age range of children (see, e.g., Redmond, 2011; Wiig et al., 2013). Repetition tasks are comprehensive, jointly measuring language recognition and production (Polišenská et al., 2015), are widely used to assess language, and have low task demands. In our task, children repeated high- and low-frequency four-word phrases in quiet and in speech-shaped background noise. Extending measures used in Bannard and Matthews (2008) and Kueser and Leonard (2020), we assessed production accuracy as well as reaction time and phrase duration.

Method

Participants

Forty children with typical language development aged 3;11 through 12;11 (mean age = 7.86 years; SD = 2.24 years) participated in the study. All participants passed a pure-tone hearing screening, had no history of seizures or neurological impairment, and had been exposed to English from birth. They were also required to pass a language screening – either the Redmond Sentence Recall Test (Redmond, 2011) or the TILLS Student Language Scale (Nelson et al., 2016). If a participant failed a screening, they were administered the SPELT-3 (Dawson et al., 2003) to rule out developmental language disorder. We noted participants' speech sound errors (e.g., "weaf" for "leaf") during these tasks and in unstructured conversation for use in scoring later (see Scoring and Reliability below). Participants completed the study in 2024 and 2025 at Boys Town National Research

Hospital in Omaha, NE (IRB 24-09-X). One participant was removed due to ceiling performance, resulting in a final sample of 39 participants (N = 20 male, 19 female). Of these participants, 12 were Black, ten were white, nine were more than one race, three were Asian, one was Native Hawaiian or other Pacific Islander, and four declined to report race. Thirty-six participants were not Hispanic/Latino, one was Hispanic/Latino, and two declined to report ethnicity. Information about dialect was not collected, but, as noted below, the phrases used in the study were dialect neutral.

Stimuli

Stimuli selection The stimuli consisted of 48 four-word phrases. Each phrase had a three-word stem followed by a phrase-final word (e.g., "corn on the ..."). Each stem was followed by a word that created a highly frequent phrase (e.g., "corn on the cob") or an infrequent phrase (e.g., "corn on the tray"), making 24 high-frequency phrases and 24 low-frequency phrases. The frequency of the phrases was measured using the North American subset of the CHILDES corpus (MacWhinney, 2000). As these stimuli were originally developed for a study of four- and five-year-old children (Kueser & Leonard, 2020), we included transcripts in CHILDES with target children from ages one to six years. The low-frequency phrases appeared at most once in the corpus. The high-frequency phrases were among the top 1% most frequent four-word phrases in the corpus. The phrases were largely dialect-neutral with limited morphological marking (e.g., noun phrases with bare nouns: "corn on the cob"; verb phrases with bare verbs: "talk on the phone"); exceptions were three phrases with plural nouns and two phrases with verbs with progressive ending.

Control variables We collected additional information about the statistical and phonological characteristics of the phrases to control for their influence so that any effect of frequency was due to units longer than two words. Phonological information about the phrases came from the CLEARPOND calculator (Marian et al., 2012). The final words in the high- and low-frequency phrases were matched in word frequency, length in phonemes, positional segment average, biphone frequency average, phonological neighborhood density, and final bigram frequency, $p > .69$, $d < 0.12$. We statistically controlled for the effects of additional nuisance characteristics of the phrases in our models using principal components analysis (PCA). These characteristics included initial trigram frequency, initial/middle/final bigram frequency, first/second/third/fourth word frequency, average word length in phonemes, average positional segment average, and average biphone frequency average. There were 11 PCs that accounted for 100% of the variance in the phrases' nuisance statistical and phonological characteristics.

Recording and synthesis A female speaker of Midwestern American English audio recorded the phrases. Final words were added to stems in Praat so that high- and low-frequency phrase pairs used the same recordings of the stems (Boersma & Weenick, 2019). The phrases' duration was normalized to 1,675 ms. For the noise condition, speech-shaped noise was added to the phrases at a signal-to-noise ratio (SNR) of 0 dB. This SNR was selected given evidence that it is a challenging listening condition for children (e.g., McCreery & Stelmachowicz, 2011). The noise was derived from the long-term average of the frequency distribution of the phrases themselves (e.g., Hall et al., 2002). Final stimulus intensity was scaled to match that of the phrases in the quiet condition.

Procedure

Each participant heard 24 phrases in the quiet condition and 24 in the noise condition. The high- and low-frequency phrases that share a stem (e.g., "corn on the cob", "corn on the tray") were randomly assigned to the same condition (e.g., quiet). The phrases were organized into four 12-item blocks that alternated between quiet and noise. For half of the pairs of phrases sharing the same three-word stem, the high-frequency phrase appeared first, and its low-frequency counterpart appeared 12 items later. Four counterbalances were created under these constraints by varying the specific phrases that were assigned to the quiet and noise conditions and the specific order of the 12-item blocks. Across counterbalances, each phrase was heard equally often in the quiet vs. noise condition and in different positions. The participants were assigned randomly to a counterbalance.

The participants completed four practice trials at the beginning of the study. They were asked to listen closely to the practice phrases and repeat them. These initial practice trials were presented in quiet. If the child was unable to independently repeat or attempt to repeat at least three of the four practice items, the practice items were presented again until the child was able to do so. Following these practice trials, the participants heard and repeated 12 phrases in the quiet condition. The participants then completed four practice trials for the noise condition. Across these practice trials, the SNR gradually changed from +10 dB to 0 dB. Instructions were the same as for the quiet condition. Following these practice trials, the participants repeated 12 phrases in the noise condition, took a short break, and then repeated the remaining 12 phrases in the quiet condition and the remaining 12 phrases in the noise condition. Only noncontingent feedback was provided during the experimental trials.

Scoring and Reliability

Accuracy Accuracy was scored for the three-word stem that was shared between the high- and low-frequency phrases. We coded word-level errors only; if a word had phonological errors (e.g., "weaf" for "leaf") that were consistent with a child's individual phonological system, then it was transcribed as the intended target. Each repetition attempt was assessed for the following errors: (1) *No Attempt*: The child did not make an attempt (e.g., "I don't know"); (2) *Repetition error*: The entirety of a word or phrase was repeated; (3) *Deletion error*: A word was not present and no other word appeared in its place; (4) *Insertion error*: A word or other isolated phonetic material appeared at the margins of or inside the three-word stem; (5) *Substitution error*: A word or word-like utterance was produced that replaced the target word. A repetition attempt was coded as incorrect if one or more of these errors was observed and correct otherwise.

Timing Correct repetition attempts were analyzed for timing. Praat was used to mark when the recorded phrase concluded and when each word within the repetition attempt began and ended. From these data, we calculated reaction time (start of first word – end of recorded phrase) and stem duration (end of third word – start of first word).

Statistical Approach

Accuracy Accuracy data for the 3,744 included trials were analyzed using a Bayesian logistic regression model. The dependent variable was correct/incorrect on the three-word stem. Predictors were frequency (high/low), age (in years, z-scored), noise condition (quiet/noise), and their interactions. PCs for nuisance statistical and phonological properties of the phrases were included as predictors; they are omitted from the results tables for brevity. Random intercepts of participant and phrase embedded in stem were included. Maximal random slopes were included where appropriate (Barr et al., 2013). The model equation was *Accuracy ~ Frequency * Age * Noise_condition + PC_control_variables + (Frequency + Noise_condition + PC_control_variables | Participant) + (Age + Noise_condition | Stem / Phrase)*. There were 2,000 warmup samples and 8,000 post-warmup samples in four chains. Fixed effect priors were wide and uninformative normal distributions (Mean = 0, SD = 0.5).

After model fitting, strength of effects was assessed using 95% credible intervals (CIs), p_d values ("*p* direction"), and p_{ROPE} values ("proportion of posterior samples in the region of practical equivalence"). p_d values are the proportion of posterior samples above or below zero for estimates above or below zero. We characterize p_d values above .95 as showing strong evidence, values from .90 to .95 as showing weak evidence, and values below .90 as showing little evidence. The region of practical equivalence (ROPE) is defined as the range of effect sizes considered to be negligible; p_{ROPE} is the proportion of posterior samples within the negligible effect size range. We characterize p_{ROPE} values above .95 as having strong evidence that the effect is at or near zero, values between .90 and .95 as having weak evidence, and values below .90 as having little evidence. An advantage of using this measure in conjunction with p_d values is that they are sensitive to sample size; if p_d values show no positive evidence for an effect, then we can look to p_{ROPE} values for an indication about whether the lack of effect is real or more likely due to small sample size. For the accuracy model, the ROPE was defined as -0.1813 to 0.1813 for a negligible effect size (Chinn, 2000; Cohen, 1988). We used *R* 4.4.2, *brms* 2.23.0, *bayestestR* 0.17.0, and *emmeans* 1.11.2-8 (Bürkner, 2018; Lenth, 2019; Makowski et al., 2019).

Timing For the timing models, only correctly repeated phrases were included (1,968 trials). Predictors, random effects, and number of samples were the same as for the accuracy model. Priors were wide and uninformative.

Reaction time data were analyzed using a Bayesian shifted lognormal regression model. The dependent variable was reaction time to begin the repetition. Additionally, shift (i.e., the earliest allowable reaction time) was modeled using an overall intercept and a random intercept for participant to allow for variation in individual processing speed. The ROPE for this model was -0.0329 to 0.0329, corresponding to an effect of ±10 ms. For brevity, only parameters for μ (difficulty) are reported in the Results.

Stem duration data were analyzed using a Bayesian linear regression model with a normal distribution. The dependent variable was the duration of the stem. The ROPE for this model was -26.50 to 26.50, corresponding to ±0.1 * SD of the distribution of the stem duration (Cohen, 1988).

Results

Accuracy

The accuracy data for each participant are plotted in Figure 1. The Bayesian logistic regression model (Table 1) showed strong evidence for main effects of frequency, age, and noise. High-frequency phrases were more often repeated accurately, EMM [estimated marginal median on proportion scale] = .90, 95% CI: [.85, .95], compared to low-frequency phrases, EMM = .85, 95% CI: [.78, .92], p_d = .988, p_{ROPE} = .284. For each additional year in age, the odds of accurately repeating a phrase were 1.37 times higher, 95% CI: [1.15, 1.66], p_d = .999, p_{ROPE} = .007. Last, phrases in the quiet condition were repeated more accurately, EMM = .92, 95% CI: [.87, .96], than phrases in the noise condition, EMM = .83, 95% CI: [.74, .90], p_d = .999, p_{ROPE} = .038. There was no evidence for interactions between frequency, age, and noise. In fact, there was strong evidence for a near-zero effect for the interaction between frequency and noise, p_{ROPE} = .951, and the interaction between frequency, noise, and age, p_{ROPE} = .986. There was weaker evidence for a near-zero effect for the interactions between frequency and age, p_{ROPE} = .928, and between age and noise, p_{ROPE} = .940.

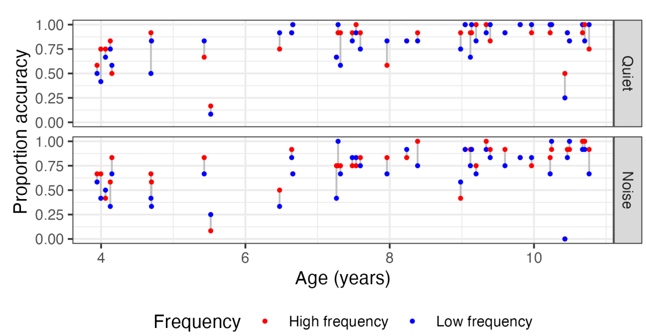

Figure 1: Proportion accuracy on stem. Lines connect data for individual participants.

Table 1: Accuracy model results.

Term	*B*	95% HDI CI	p_d	p_{ROPE}
Intercept	2.01	[1.46, 2.55]	1	0
Frequency	0.24	[0.04, 0.44]	.988	.284
Age	0.70	[0.31, 1.12]	.999	.007
Noise	0.43	[0.16, 0.70]	.999	.038
F x A	-0.07	[-0.22, 0.09]	.797	.928
F x N	-0.05	[-0.20, 0.20]	.714	.951
A x N	0.00	[-0.18, 0.20]	.518	.940
F x A x N	-0.02	[-0.16, 0.11]	.626	.986

Note. F = Frequency; A = Age; N = Noise.

Timing

The reaction time data (i.e., time to respond after hearing a phrase) for accurately repeated phrases are shown in Figure 2 for each participant. The Bayesian shifted lognormal model (Table 2) showed strong evidence for the effect of frequency.

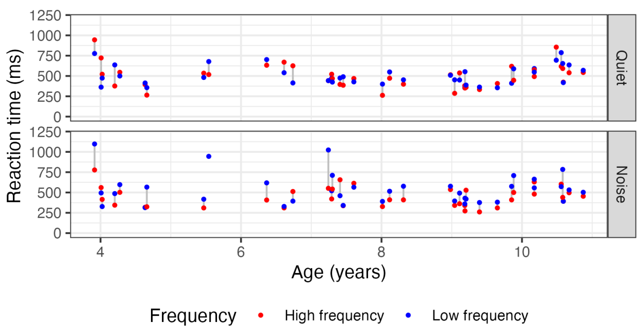

Figure 2: Reaction time. Lines connect data for individual participants.

Table 2: Reaction time model results.

Term	*B*	95% HDI CI	p_d	p_{ROPE}
Intercept	5.70	[5.57, 5.84]	1	0
Frequency	-0.08	[-0.12, -0.04]	.999	.020
Age	-0.04	[-0.15, 0.06]]	.803	.342
Noise	0.02	[-0.04, 0.07]	.742	.669
F x A	0.01	[-0.02, 0.05]	.755	.869
F x N	0.03	[-0.01, 0.06]	.947	.626
A x N	0.01	[-0.04, 0.06]	.686	.795
F x A x N	-0.01	[-0.04, 0.02]	.774	.905

Note. F = Frequency; A = Age; N = Noise.

High-frequency phrases were repeated more quickly, EMM = 437 ms, 95% CI: [394 ms, 488 ms], than low-frequency phrases, EMM = 495 ms, 95% CI: [444 ms, 552 ms], p_d = .999, p_{ROPE} = .020. There was moderate evidence for an interaction between frequency and noise condition (Figure 3) such that the difference between high- and low-frequency phrases in the quiet condition, $EMM_{Low\text{-}High}$ = 37.40 ms, 95% CI: [-1.2 ms, 76.90 ms], was smaller than the difference in the noise condition, $EMM_{Low\text{-}High}$ = 74.70 ms, 95% CI: [-34.30 ms, 118.70 ms], p_d = .947, p_{ROPE} = .626. This difference appeared to be more driven by differences in the high-frequency phrases across conditions, $EMM_{Noise\text{-}Quiet}$ = -29.72 ms, 95% CI: [-72.90, 13.60], compared to change in the low-frequency phrases, $EMM_{Noise\text{-}Quiet}$ = 8.52 ms, 95% CI: [-41.70, 60.20]. There was no evidence for other effects and weak evidence that the interaction between frequency, age, and noise had near-zero effect, p_{ROPE} = .905.

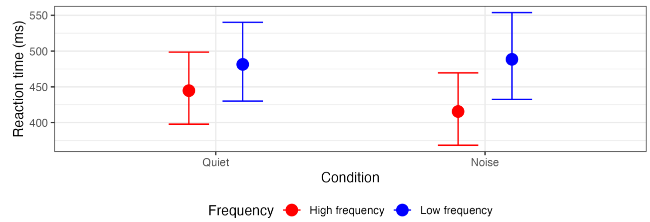

Figure 3: Reaction time model predictions for frequency x noise condition interaction. Error bars are 95% CIs.

The stem duration data for accurately repeated phrases are presented in Figure 4 for each participant. The Bayesian linear regression model (Table 3) showed strong evidence that all predictors and their interactions had near-zero effect, p_d < .577, p_{ROPE} = 1.

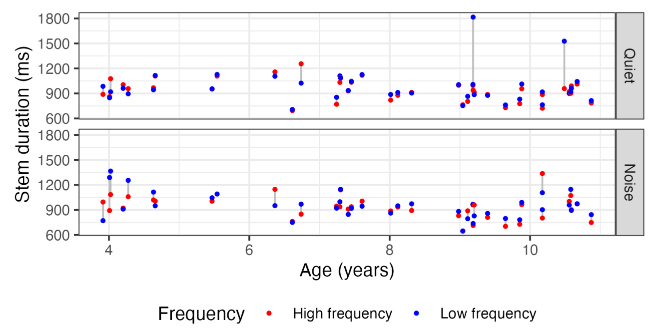

Figure 4: Stem duration. Lines connect data for individual participants.

Table 3: Stem duration model results.

Term	*B*	95% HDI CI	p_d	p_{ROPE}
Intercept	945.66	[0.89, 1.00]	1	0
Frequency	-0.04	[-1.05, 0.93]	.535	1
Age	-0.03	[-1.00, 0.93]	.529	1
Noise	-0.01	[-0.97, 0.96]	.504	1
F x A	-0.02	[-0.98, 0.97]	.517	1
F x N	-0.01	[-0.98, 0.99]	.509	1
A x N	-0.00	[-0.96, 1.00]	.503	1
F x A x N	-0.10	[-1.06, 0.89]	.577	1

Note. F = Frequency; A = Age; N = Noise.

Discussion

Prior work had shown that word frequency aids word processing across development and in background noise. Although phrase frequency had also been shown to aid phrase processing, it was not known how the effect of phrase frequency developed over childhood and interacted with background noise. In the current study, children's phrase processing was found to have both similarities to and differences from prior work.

Among the similarities to prior work, we first found a facilitatory effect of phrase frequency on repetition accuracy and speed, consistent with prior studies that used a similar design (Bannard & Matthews, 2008; Kueser & Leonard, 2020). Second, we found a developmental trend such that older children had more success in repetition than younger children, again, a finding consistent with prior work on improvements in repetition ability over development (e.g., Dollaghan & Campbell, 1998). Third, we found an effect of listening condition, with phrases in quiet being repeated better than phrases in background noise. Overall, these findings indicate that children have better repetition performance with higher-frequency phrases and in quiet, and their performance improves over development.

Other patterns, however, offered new insights that differed from prior work. First, we did not find an interaction between phrase frequency and age in the accuracy or the timing data. For the phrases under study, the high-frequency advantage was about the same across the age range. This suggests that phrase frequency information is learned and used quickly across childhood and that the ability to learn this information seems relatively consistent across development.

However, factors potentially complicating this conclusion included that (1) the phrases were simple and short, which may have limited the potential effect of frequency as all the phrases were relatively easy to repeat, (2) the difference between high and low frequency in the corpus was very large, which may have masked intermediate effects that require more sophisticated language knowledge only present in older children, (3) while we eliminated the child who scored at ceiling overall, some children reached ceiling in individual conditions, particularly at older ages for high-frequency phrases, and this could have attenuated larger frequency effects in the older group; and (4) the sample size was relatively small, though we note that the p_{ROPE} values indicating a lack of an effect are sensitive to sample size. Future work should use longer phrases or examine smaller frequency differences to address these limitations.

There was mixed evidence for an interaction between phrase frequency and noise. In the accuracy data, this interaction was not observed. This was somewhat unexpected, given that work on word frequency shows that frequency is likely to have a larger effect in background noise vs. quiet. Again, this could be because the high- and low-frequency phrases were short, and so they were not difficult enough to show a strong effect. Additionally, perhaps more challenging listening conditions would show greater frequency effects since we observed some ceiling performance for the older children. Some work suggests that effects of informational masking (i.e., use of competing background speech) differs from the energetic noise masking used here (Buss et al., 2019; Erickson & Newman, 2017). Future work may use strategically designed background babble to further test the effects of phrase frequency given informational masking. Additionally, examination of the specific errors made in each condition, including for the final word, could reveal how language knowledge was used in noise.

In contrast, there was some positive evidence for a frequency by noise interaction in the reaction time data. We found that the latency advantage for high- vs. low-frequency phrases was greater for the noise compared to the quiet condition. On one interpretation, this pattern suggests that the process of "reconstructing" the phrase after hearing it was more difficult in noise compared to quiet. However, we would expect this to primarily affect low-frequency phrases, with greater latencies for low-frequency phrases in noise vs. quiet. In fact, we found that the interaction was driven by greater latencies for high-frequency phrases in quiet vs. noise. It is not clear why this happened. One possibility is that aspects of the experimental design caused greater change in high-frequency phrases compared to low-frequency phrases. For example, given that the first noise block always occurred after the first quiet block for all children, the children had more practice with the task when it came to the noise condition. On this explanation, we might posit that high-frequency phrases simply experienced more practice effects than low-frequency phrases. Another possibility is that the noise condition promoted greater attention, being more distinct than usual listening conditions. The same logic would apply, however: the high-frequency phrases experienced more advantage from greater attention in the noise vs. quiet conditions compared to the low-frequency phrases.

The results also have implications for models of language development. First, the results show that phrase frequency may affect language learning even in noisy listening contexts that children often encounter in real life. While noise did impact overall comprehension, it did not do so in a way that negated the possibility of recognizing high-frequency phrases. Second, the results connect with longstanding debates about the nature of linguistic frequency – is this information stored like a list of numbers or is it epiphenomenal, a product of entrenched processing (e.g., Schmid, 2010)? The patterns here are consistent with the idea that phrase frequency effects reflect well-worn, entrenched processing pathways rather than a list of stored numbers that can be called up at will. The near-zero interaction between frequency and noise in the accuracy data supports the idea that phrase frequency is not a piece of numeric information that can be preferentially tuned into in background noise. Instead, phrase frequency represents bias in the language system and so is used equivalently in both background noise and quiet.

In conclusion, across typical development, phrase frequency aids language processing in quiet and in background noise. Evidence for preferential use of frequency information in background noise was mixed, with the strongest evidence showing that it was not preferentially used to overcome background noise.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number P20GM109023.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Ambridge, B. (2019). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, 51.

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(02), 239–273. https://doi.org/10.1017/S030500091400049X

Archibald, L. M. D., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*, 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science*, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Boersma, P., & Weenick, D. (2019). *Praat: Doing phonetics by computer* (Version 6.0.49) [Computer software]. http://www.praat.org/

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, *74*(1), 1–15. https://doi.org/10.1037/h0024206

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., & Grose, J. H. (2019). Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type. *Ear and Hearing*, *40*, 1117–1126.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*(22), 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22%253C3127::AID-SIM784%253E3.0.CO;2-M

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*(3), 542–551. https://doi.org/10.1111/tops.12274

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, *12*(4), 425–434. https://doi.org/10.1177/014662168801200410

Cooke, M., García Lecumberri, M. L., Barker, J., & Marxer, R. (2019). Lexical frequency effects in English and Spanish word misperceptions. *The Journal of the Acoustical Society of America*, *145*(2), EL136–EL141. https://doi.org/10.1121/1.5090196

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing*, *37*, 55–63.

Davies, R. A. I., Arnell, R., Birchenough, J. M. H., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1298–1338. https://doi.org/10.1037/xlm0000366

Dawson, J., Stout, C. E., & Eyer, J. (2003). *Structured Photographic Expressive Language Test: Third Edition*. Janelle Publications.

DeDe, G., & Flax, J. K. (2016). Language comprehension in aging. In H. H. Wright (Ed.), *Cognition, Language and Aging*. John Benjamins Publishing Company. https://doi.org/10.1075/z.200

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, *26*(5), 451–457. https://doi.org/10.1177/0963721417709087

Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, *23*(2), 159–165. https://doi.org/10.1097/00003446-200204000-00008

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America*, *29*(2), 296–305. https://doi.org/10.1121/1.1908862

Krull, V., Choi, S., Kirk, K. I., Prusick, L., & French, B. (2010). Lexical effects on spoken-word recognition in children with normal hearing. *Ear & Hearing*, *31*(1), 102–114. https://doi.org/10.1097/AUD.0b013e3181b7892f

Kueser, J. B., & Leonard, L. B. (2020). The effects of frequency and predictability on repetition in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *63*(4), 1165–1180.

Lenth, R. (2019). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Vol. 2. The database*. Lawrence Erlbaum Associates.

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science*, *34*(3), 465–488. https://doi.org/10.1111/j.1551-6709.2009.01091.x

McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science*, *9*(3), 637–652. https://doi.org/10.1111/tops.12258

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America*, *130*(6), 4070–4081. https://doi.org/10.1121/1.3658476

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. https://doi.org/10.1037/h0027366

Nelson, N. W., Plante, E., Helm-Estabrooks, N., & Hotz, G. (2016). *Test of Integrated Language and Literacy Skills (TILLS)*. Paul H. Brookes Publishing Co.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: What does the task measure? *International Journal of Language and Communication Disorders*, *50*(1), 106–118. https://doi.org/10.1111/1460-6984.12126

Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *The Journal of the Acoustical Society of America*, *32*(4), 454–457. https://doi.org/10.1121/1.1908097

Redmond, S. M. (2011). *Redmond Sentence Recall Manual*.

Schmid, H.-J. (2010). Does frequency in text instantiate entrenchment in the cognitive system? In D. Glynn & K. Fischer (Eds.), *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches* (pp. 101–133). De Gruyter Mouton. https://doi.org/10.1515/9783110226423

Strauß, A., Wu, T., McQueen, J. M., Scharenborg, O., & Hintz, F. (2022). The differential roles of lexical and sublexical processing during spoken-word recognition in clear and in noise. *Cortex*, *151*, 70–88. https://doi.org/10.1016/j.cortex.2022.02.011

Van Engen, K. J., Dey, A., Runge, N., Spehar, B., Sommers, M. S., & Peelle, J. E. (2020). Effects of age, word frequency, and noise on the time course of spoken word recognition. *Collabra: Psychology*, *6*(1), 17247. https://doi.org/10.1525/collabra.17247

Walker, E. A., Sapp, C., Oleson, J. J., & McCreery, R. W. (2019). Longitudinal speech recognition in noise in children: Effects of hearing status and vocabulary. *Frontiers in Psychology*, *10*, 2421. https://doi.org/10.3389/fpsyg.2019.02421

Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, *118*(5), 3164–3176. https://doi.org/10.1121/1.2082567

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. Pearson.